Immunogenetics and polymorphism in a natural population of field voles (*Microtus agrestis*)

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Andrew Keith Turner

August 2010

Abstract

Most of our understanding of immunity has been gained through studies of humans or laboratory rodents. However, such studies do not allow the immune system to be studied in the ecological context in which it has evolved and, as such, they provide a poor model for studying the variation in infectious disease resistance and immune function observed in natural settings. Studies of natural populations have provided fresh insights into the evolution and phenotypic consequences of immunogenetic variation, but have thus far concentrated almost exclusively on genes of the major histocompatibility complex (MHC). As these genes form only a fraction of the vertebrate immune repertoire, there is a need to broaden research in natural populations to include non-MHC genes, in order to gain a more comprehensive understanding of natural selection and immunity.

In this thesis, the genetic diversity of a range of non-MHC immune genes was examined in a natural population of field voles (*Microtus agrestis* L.) in Kielder Forest, UK, which are subject to infection by a range of pathogens. Cytokine genes were the primary focus of this study as they play a central role in regulating the immune response but have rarely been studied in wild species. I examined the hypothesis that, as cytokines are crucial to immunity, variation within these genes may be under selection within populations and between species, and may mediate phenotypic differences between individuals in parasite resistance and immune function.

Coding regions from nine cytokine and three other non-MHC vole genes were sequenced, yielding 6.6 Kb of sequence data and 26 SNPs (1 per 255 bp). Three cytokine genes (*Il1b*, *Il2*, and *Tnf*) exhibited patterns of polymorphism consistent with balancing selection maintaining genetic diversity, including an excess of intermediate frequency mutations and more even allele frequencies than one would expect under neutrality.

Polymorphism within *Il1b* and *Il2* was also consistently associated with variation in parasite resistance, providing evidence that pathogens are the selective force driving the maintenance of genetic diversity at these loci. In addition, *Il1b* and *Il2* exhibited repeated associations with variation in host immune phenotype, while the *Il12b* gene was associated both with variation in pathogen resistance and with altered expression levels of *Il1b* and *Il2*. Variation in immune function, mediated through the cytokine network, is therefore likely to contribute to parasite resistance in the field vole.

This work is the first to show that variation in cytokine genes of a natural population can be maintained by selection, and that this variation can lead to phenotypic variation in parasite resistance and immune function. More broadly, this thesis demonstrates that wild rodents are an excellent model to help us bridge the gap in our understanding between the mechanistic insights gained through studies on laboratory rodents and the variation in infectious disease susceptibility and immune function observed in nature.

Acknowledgments

Firstly I'd like to thank my supervisors Dr Steve Paterson and Prof. Mike Begon for their help and support during my PhD. They have trusted me throughout to get on with my own work and make my thesis my own, but have always been available for advice or a friendly kick in the right direction when necessary. Most of all I'm extremely grateful to both of them for retaining their optimism when I thought my PhD was buggered. Thanks also to NERC for funding my project and to the many post-docs who have offered advice, notably Helen O'Meara and Joe Jackson at either end of my PhD.

I am extremely lucky to have met so large a group of friends at Liverpool that there isn't room to write them all here. They have helped me to enjoy studying for a doctorate more than I probably should have been allowed, and stuck by me as I've turned into someone who can no longer handle mid-week drinks and can get a hangover from a curry. Special thanks must go to Becci, who has taken the brunt of my stress and moaning, for her support.

Finally, thank you to my family for their constant support throughout my PhD and always. I'm very fortunate to have a family who are so proud of me, despite not actually properly knowing what I do…

And Dad, no, I haven't cured cancer yet.

Contents

List of tables

List of figures

Glossary of abbreviations

H_d Haplotype diversity

μ Mutation rate

AIC Akaike Information Criterion

cDNA Complementary DNA

CDS Coding sequence

DNA Deoxyribonucleic acid

dNTP Deoxyribonucelotide

F_{IS} Inbreeding coefficient; a measure of the extent of genetic inbreeding within subpopulations

***Foxp3* / FOXP3** Forkhead box P3

F_{ST} Fixation index; a measure of genetic differentiation between subpopulations

***Gata3* / GATA3** GATA-binding protein 3

gDNA Genomic DNA

GLM Generalised linear model

GLMM Generalised linear mixed model

H_E Expected heterozygosity under Hardy-Weinberg equilibrium

HKLM Heat-killed *Listeria monocytogenes*

H_o Observed heterozygosity

H_S Heterozygosity within a subpopulation

H_T Heterozygosity of the total population

HWE Hardy-Weinberg equilibrium

***Ifng* / IFN-γ** Interferon gamma

***Il10* / IL-10** Interleukin 10

***Il12a* / IL-12α** Interleukin 12, alpha

***Il12b* / IL-12β** Interleukin 12, beta

***Il18* / IL-18** Interleukin 18

***Il1a* / IL-1α** Interleukin 1, alpha

***Il1b* / IL-1β** Interleukin 1, beta

***Il2* / IL-2** Interleukin 2

***Il4* / IL-4** Interleukin 4

***Il5* / IL-5** Interleukin 5

***Irf5* / IRF-5** Interferon regulatory factor 5

LD Linkage disequilibrium

LM General linear model

LPS Bacterial lipopolysaccheride

MAF Minor allele frequency

MHC Major histocompatibility complex

mRNA Messenger RNA

MYA Million years ago

N_e Effective population size

PAMP Pathogen-associated molecular pattern

PCA Principal components analysis

PCR Polymerase chain reaction

PHA Phytohaemagglutinin

PRR Pattern-recognition receptor

RNA Ribonucleic acid

***Slc11a1* / SLC11A1** Solute carrier family 11 (proton-coupled divalent metal ion transporters), member 1

SNP Single nucleotide polymorphism

T_a Annealing temperature

***Tbx21* / TBX21** T-box 21

***Tgfb1* / TGF-β1** Transforming growth factor, beta 1

Th(1,2,17) T-helper cell (type 1, 2 or 17)

***Tlr2* / TLR2** Toll-like receptor 2

***Tlr4* / TLR4** Toll-like receptor 4

T_m Melting temperature

***Tnf* / TNF** Tumour necrosis factor

Treg Regulatory T-helper cell

η Total number of mutations within a gene sequence

η_e Total number of mutations in external branches of a gene genealogy

θ Neutral population parameter; expected nucleotide variation for a neutrally evolving diploid population is $\theta = 4N_e\mu$

θ_H Fay and Wu's estimator of θ, weighted by the homozygosity of the derived variants

θ_W Watterson's estimator of θ

π Nucleotide diversity (average number of nucleotide substitutions per site)

Chapter 1

Introduction

1.1 INFECTIOUS DISEASE AND EVOLUTION

All free-living populations are subject to infection by a wide range of parasites that, by definition, can reduce the fitness of their hosts. Individuals who are more capable of resisting infection are more likely to survive and reproduce and, as such, parasites have always acted as a strong selective force on host populations (Barreiro and Quintana-Murci 2010). As a result, many organisms have evolved sophisticated innate and adaptive immune systems, particularly so in the case of vertebrates, to resist or control such infections (Kimbrell and Beutler 2001; Flajnik and Du Pasquier 2004; Cooper and Alder 2006). The resulting selective pressures of the host immune response in turn drive pathogenic organisms to evolve new mechanisms of avoiding recognition by the immune system and of infecting new hosts (Slev and Potts 2002). As first highlighted by Haldane (1949), these co-evolutionary dynamics between hosts and parasites can act as the selective force to drive the genetic divergence between species, or to maintain polymorphism within populations.

Since Haldane, a large body of empirical evidence has supported his hypothesis that parasites are a major selective force in shaping host genetic diversity (Sommer 2005; Barreiro and Quintana-Murci 2010). Genes of the immune system may be particularly likely to experience strong parasite-driven selective pressures, as these are the genes that interact with the myriad of pathogens that constantly challenge free-living organisms. Several studies have demonstrated that the rate of adaptive evolution is higher at immune loci than other classes of gene in comparisons between *Drosophila* species (Schlenke and Begun 2003; Sackton *et al.* 2007; Obbard *et al.* 2009), between humans and mice (Huang *et al.* 2004) and in fish (Tonteri *et al.* 2010). However, none of these studies have attempted to examine

the phenotypic consequences of this genetic diversity (i.e. how genetic variation relates to parasite resistance or immune function), which is essential to show that parasites are the selective force driving the evolution of immune genes. More substantial evidence for pathogen-driven selection has come from studying genes of the major histocompatibility complex (MHC), an important and highly polymorphic family of immune genes involved in antigen recognition and presentation. The MHC exhibits astonishingly high levels of genetic polymorphism which has been maintained through strong balancing selection (Piertney and Oliver 2006). Evidence that the high levels of polymorphism seen within MHC loci are driven by parasites comes from numerous studies in human, laboratory and natural populations, where specific associations between MHC variation and resistance to pathogens have been found (Else and Wakelin 1988; Else and Wakelin 1989; Keymer *et al.* 1990; Hill 1998; Paterson *et al.* 1998; Reusch *et al.* 2001; Kurtz *et al.* 2004; Piertney and Oliver 2006). Such studies demonstrate the potential for parasites to shape the genetic diversity of immune genes. However, the genes of the MHC account for only a fraction of the variation among individuals in pathogen susceptibility (Jepson *et al.* 1997) and there are therefore many other genes which contribute both to the immune response and to variation in infectious disease susceptibility. Despite this, studies on the immunogenetics of natural populations have concentrated almost exclusively on the MHC. In order to broaden our understanding of the effect natural selection has on immunity, genetic diversity and fitness, there is a need to examine a much wider range of immune-function genes within natural populations (Acevedo-Whitehouse and Cunningham 2006).

1.2 CYTOKINES AND THE IMMUNE RESPONSE

This thesis will expand research on wildlife immunogenetics away from the MHC and primarily focus on genes encoding cytokines, which are signalling molecules fundamental to the immune response. Cytokines regulate and modulate almost every aspect of vertebrate immunity including inflammatory, humoral and cell-mediated immune responses, bone marrow differentiation, antigen presentation and adhesion molecule expression (Borish and Steinke 2003). They are secreted from

many cell types in response to a variety of stimuli. Cytokines have their effect by binding to their respective receptor on a cell surface, which results in the activation of signal transduction and second messenger pathways within a cell (Smith and Humphries 2009). These molecules usually act in an autocrine or paracrine manner (acting upon the cell that secreted them, or nearby cells, respectively) and are extremely potent, acting at picomolar or sometimes femtomolar levels (Balkwill 2002).

Cytokines represent a highly diverse group of molecules that form a hugely complex network of interacting proteins These molecules exert a wide range of actions on a great many immune cell types (Figure 1.1). Many individual cytokines are pleiotropic and multi-functional and, although there are hundreds of individual cytokine genes, the mature proteins often exhibit extraordinary levels of redundancy, in that many of these molecules have overlapping actions (Ozaki and Leonard 2002).

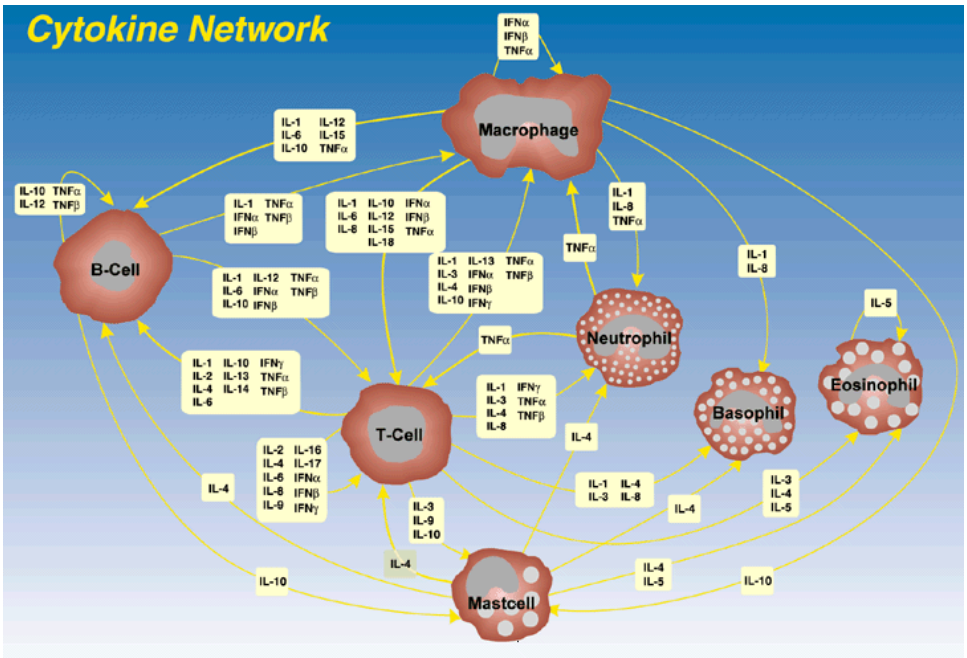

Figure 1.1 A much-simplified representation of the cytokine network.
Diagram from *The Scientist* 2000, **14**(7):36

The type and extent of immune response elicited by the host is largely determined by which cytokines are produced in response to a given immune challenge. There are several distinctive effector arms of the immune response, each driven by particular $CD4^+$ T-helper (Th) cells, which are central to adaptive immunity. There are four broad Th-cell phenotypes, Th1, Th2, Treg and Th17, each of which can be identified by the cytokines they produce. Although in reality there may be considerable overlap between the Th phenotypes, in general terms each is associated with a characteristic function.

Polarization of Th1 cells from naïve Th0 cells is directed through the action of the cytokine interleukin (IL)-12. Mature Th1 cells express the transcription factor *Tbx21* and produce interferon (IFN)-γ, which promotes immune responses (such as phagocytosis or oxidative bursts) directed against intracellular microparasites (including viruses, bacteria, fungi and protozoa) (Mosmann *et al.* 1986; Abbas *et al.* 1996; Graham *et al.* 2007). In contrast, Th2 cells are associated with immune responses against macroparasites, primarily helminths. IL-4 is the major driver of Th2 differentiation. Mature Th2 cells express the transcription factor *Gata3* and produce the cytokines IL-4, IL-5, IL-9 and IL-13 (Bradley and Jackson 2008). Traditionally, a simple counterbalance and cross-regulation between the Th1 and Th2 arms was thought to be the primary mechanism for controlling Th responses (Mosmann *et al.* 1986), but it is now recognised that regulatory T-helper cells (Tregs) exert a predominant suppressive effect on all immune effector cell types (Bluestone and Abbas 2003). Treg cells typically express the transcription factor *Foxp3* (Hori *et al.* 2003) and the activation marker CD25 and are both induced by, and producers of, the cytokines IL-10 and transforming growth factor (TGF)-β (Bradley and Jackson 2008). Finally, Th17 cells represent a recently described category of effector T-cells thought to be important in the clearance of extracellular microparasites. Development of these cells also depends on TGF-β, and they are characterised by the production of IL-17, IL-22 and the expression of the transcription factor *ROR-γt* (Weaver *et al.* 2006; Bradley and Jackson 2008). Quantification of cytokine production is relatively straightforward and, coupled with their central role in immunity, cytokines therefore offer a functionally relevant,

measurable way to simplify and characterise the multifaceted vertebrate immune response (Graham *et al.* 2007).

1.3 FUNCTIONAL CONSEQUENCES OF CYTOKINE POLYMORPHISM

Given the importance of cytokines in immunity, it is perhaps not surprising that genetic variation leading to structural or quantitative expression changes may be associated with altered susceptibility to disease. There are now many thousands of studies highlighting associations between variation within cytokine genes and infectious disease (Hill 2001a; Ollier 2004; Hollegaard and Bidwell 2006), and evidence that parasites can act as a major selective force shaping the genetic diversity of some cytokines (Fumagalli *et al.* 2009). However, almost all of our understanding of the phenotypic consequences of cytokine polymorphism has been gained from studies of humans or laboratory model species. Such studies are however often divorced from an ecological context; laboratory experiments usually study inbred, domesticated hosts which are not representative of the natural genetic variation observed in the wild. Furthermore, such research has often studied the outcome of infection by a single parasite in well-fed hosts, kept in relatively sterile and stable conditions. However, this is not the context within which infection normally occurs (Grenfell *et al.* 2002). In the wild, multiple concurrent infections, heterogeneous environments and sub-optimal nourishment are the norm. There is then currently a wide gap between our understanding of the mechanisms of immunity in laboratory models and the application of this knowledge to studies of natural populations (Bradley and Jackson 2008). For a better understanding of how host/parasite interactions shape the vertebrate immune response and the role of natural selection, ecology must inform immunology (Grenfell *et al.* 2002). There is, however, still little known about the role of selection on shaping cytokine genetic diversity or the phenotypic effects of this variation in natural populations. To my knowledge, only one example has been published on cytokine polymorphism and disease resistance in a natural population: Coltman *et al.* (2001) found an association between an intronic microsatellite within the *Ifng* gene and resistance to gastrointestinal nematodes in wild Soay sheep. However, the overwhelming

majority of immunogenetic studies in wildlife have concentred only on genes of the MHC; there is therefore a need for the broader research on cytokines and other non-MHC genes undertaken on traditional model species to be transferred to wild populations (Acevedo-Whitehouse and Cunningham 2006). It is also important that any such research should be comprehensive and incorporate both evolutionary and functional aspects. Integrating approaches from several functional (DNA sequence polymorphism, mRNA expression, phenotypic variation) and biological (individual, population, species) levels will both increase the reliability of the findings and allow a more complete understanding of the role and consequences of cytokine molecular variation in ecological and evolutionary processes (Vasemägi and Primmer 2005; Dalziel *et al.* 2009).

1.4 STUDY SPECIES AND SITE

The species chosen for study was the field vole, *Microtus agrestis* L. (Figure 1.1) within Kielder Forest, a man-made spruce forest located on the border of England and Scotland (55º13'N, 2º33'W) (Figure 1.2) The field vole is believed to be the most abundant of terrestrial British mammals with an estimated UK population size of 75,000,000 (Harris *et al.* 1995), and has a geographic range that extends throughout much of the northern region of Western Europe (Corbet and Harris 1991). In Kielder, field voles inhabit grassy clear-cuts colonized primarily by *Deschampsia caespitosa, Agrostis capillaries* and *Juncus effuses* (Figure 1.2) and are completely absent from forested areas (Burthe *et al.* 2008a; Jackson *et al.* in review).

Figure 2.2 The field vole, *Microtus agrestis*. Photo is author's own.

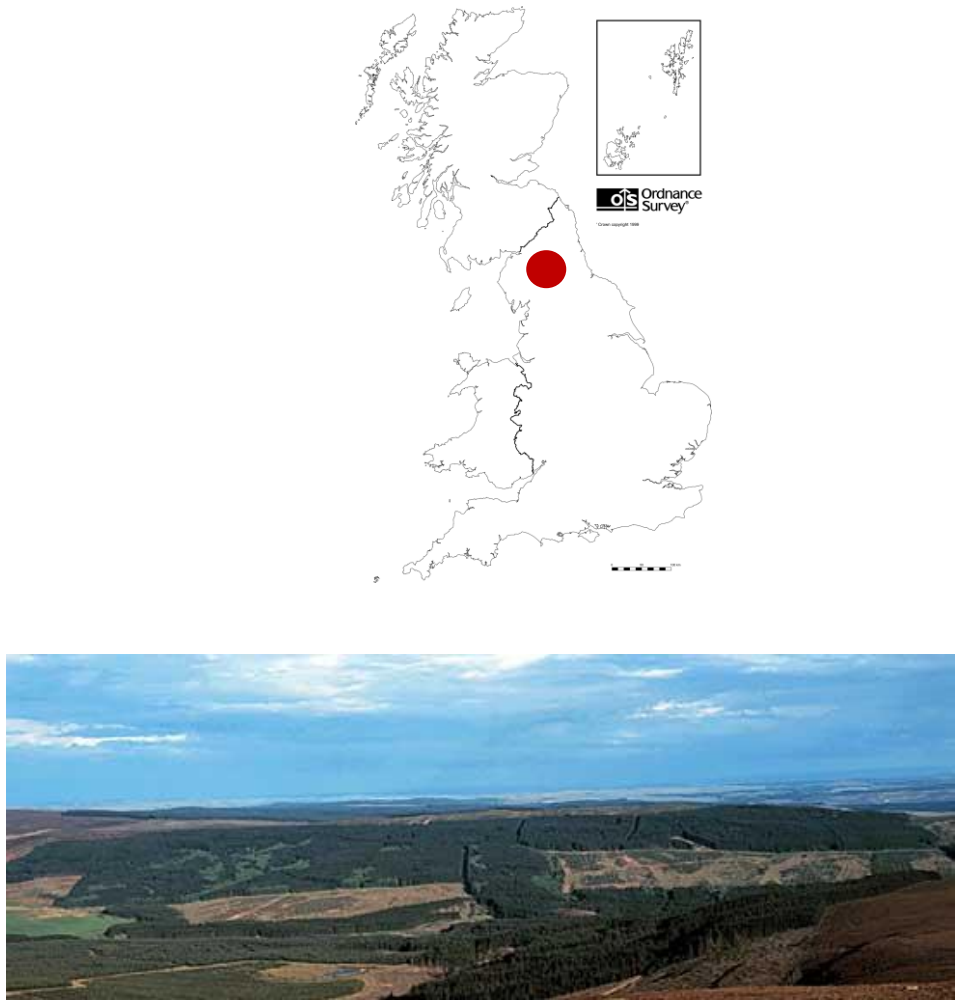

Figure 1.3 (Top) location of Kielder Forest and (bottom) view of Kielder showing several deforested clear-cut areas, inhabitable by field voles.

Kielder Forest field voles represent a particularly attractive system for the study of immunogenetics in natural populations. The Kielder voles are part of an extremely well-studied system with a long and varied background of research into population ecology and pathogen dynamics (e.g. Elton *et al.* 1935; Lambin *et al.* 2000; Cavanagh *et al.* 2002; Cavanagh *et al.* 2004; Burthe *et al.* 2006; Telfer *et al.* 2007a; Burthe *et al.* 2008a; Burthe *et al.* 2008b). They are related to laboratory species with well-characterised genomes, with an estimated divergence time of ~24.7 mya between voles and both *Mus musculus* and *Rattus norvegicus* (Steppan *et al.* 2004). This allows the extensive genomic data gathered through analysis of lab species to be both relevant and easily transferred to this natural population and also greatly facilitates the identification of orthologous immune genes in the field vole. Most

importantly, studying immunogenetics in natural populations means that many of the constraints associated with studying laboratory organisms, such as inbreeding and the lack of ecological validity, will not apply and data can therefore be analysed in a true ecological and evolutionary context. Field voles are subject to infection by a wide range of viruses, bacteria, protozoa, helminths and arthropods (Cavanagh *et al.* 2002; Cavanagh *et al.* 2004; Smith *et al.* 2005; Bown *et al.* 2006; Telfer *et al.* 2007a; Telfer *et al.* 2008). Many of these parasites are zoonotic or related to pathogens of medical or veterinary importance and may therefore provide fresh insights into the mechanisms of disease resistance in humans or livestock. Well-studied natural rodent populations such as the Kielder field voles therefore represent a system that can bridge the gap between the mechanistic understanding of infectious disease gained through studies of laboratory rodents, and the observed patterns of immunogenetic diversity and variation in disease susceptibility in nature.

1.5 AIMS

There is a need to expand research on wildlife immunogenetics away from concentrating solely on the MHC in order to improve our understanding of the mechanisms of immunity and the role of natural selection in shaping patterns of genetic diversity (Acevedo-Whitehouse and Cunningham 2006; Charbonnel *et al.* 2006). A better understanding of the link between genetics and infectious disease in natural populations will increase our knowledge of the mechanisms behind host-parasite interactions, disease pathogenesis, parasite transmission and epidemiology. In turn, this may inform decisions on the control of zoonotic or emerging diseases, conservation of threatened species and may ultimately lead to new therapeutics (Cleaveland *et al.* 2002; Acevedo-Whitehouse and Cunningham 2006; Charbonnel *et al.* 2006; Smith *et al.* 2009).

The work in this thesis will describe the genetic diversity within a range of non-MHC immune genes in a natural population of field voles and examine both the evolutionary pressures that have shaped this diversity and how this genetic variation relates to phenotypic variation in parasite resistance and immune function:

Chapter 2 - SNP discovery and sequence diversity within field vole immune genes

Coding regions of a range of field vole immunity genes were sequenced. Single nucleotide polymorphisms discovered within these genes were described and characterised. These data were used to estimate measures of sequence-level genetic diversity and several molecular evolution analyses were performed to test for evidence of natural selection.

Chapter 3 - Population genetics and natural selection at immune gene loci

Population-level genetic diversity was examined by genotyping several hundred individuals at numerous immune SNP loci. Estimates of allele frequencies across several subpopulations were used to describe patterns of Hardy-Weinberg and linkage equilibria, population structure and to accrue further evidence for natural selection acting on several immune genes.

Chapter 4 - Immune gene polymorphism and parasite resistance

The functional relevance of observed genetic polymorphism and evidence for pathogen-driven selection was tested for by examining associations between genetic diversity and resistance to several species of parasite. Linear models and related statistical techniques were used to disentangle genetic from non-genetic effects by accounting for confounding intrinsic and extrinsic variables.

Chapter 5 – Immunogenetic diversity and variation in immune phenotype

In order to explore the possibility that variation in immune phenotype is the biological mechanism underlying variation in parasite resistance, associations between immunogenetic polymorphism and variation in immune function were examined. Immune phenotype was measured through analysis of mRNA expression of several cytokines and other immune genes, and statistical models were again used to account for confounding environmental variables in tests of genetic association.

Chapter 2

SNP discovery and sequence diversity within field vole immune genes

2.1 INTRODUCTION

The characterisation of genetic diversity and an understanding of the processes responsible for its production and maintenance are of fundamental importance to the fields of population genetics, molecular ecology and evolutionary biology. A large proportion - approximately 5% - of the genome of most mammals is dedicated to immunity (Trowsdale and Parham 2004) and diversity within these genes may be of particular importance in relation to fitness; genetic variation at many immune loci has been associated with variation in resistance to many infectious and non-communicable diseases (Hill 2001b; Trowsdale and Parham 2004; Hill 2006).

The study of immunogenetic variation in non-human, natural populations has often been overlooked and yet a better understanding of diversity within immune genes of natural populations has great potential in increasing our understanding of host-parasite interactions, pathogen transmission and epidemiology, zoonotic and emerging diseases, ecology and evolution of infectious diseases and conservation management (Cleaveland *et al.* 2002; Acevedo-Whitehouse and Cunningham 2006; Charbonnel *et al.* 2006). Genes of the major histocompatibility complex (MHC) represent one branch of immunogenetics which has been extensively studied in wildlife and has become a paradigm of how parasite-driven natural selection can act to maintain diversity in natural populations (Piertney and Oliver 2006). However, in order to broaden our understanding of both the mechanisms of immunity and how selection shapes genetic variation there is now a need to move research away from the MHC to other immune-function genes in natural populations (Grenfell *et al.* 2002; Acevedo-Whitehouse and Cunningham 2006).

A small number of studies have attempted to characterise the diversity within non-MHC immune genes of non-human populations, but in the main these have focussed on domesticated animals such as chicken (Downing *et al.* 2009) or cattle (Heaton *et al.* 2001; Freeman *et al.* 2008). There are a very few examples of publications that have examined non-MHC immune gene diversity in natural populations of mammals. Coltman *et al*. (2001) found an association between polymorphism at an *Ifng*-linked microsatellite and resistance to gastrointestinal nematodes in free-living Soay sheep, while Worley *et al*. (2006) examined diversity and evidence for selection both for microsatellites and at the DNA sequence level for the genes *Ifng* and *Nramp1* (now *Slc11a1*) in wild thinhorn sheep.

The complexity of the vertebrate immune response and the narrow range of genes studied in the wild thus far means that there is a need to characterise the genetic diversity within a much wider range of immune genes in natural populations. Wild rodents represent especially promising models for expanding our knowledge of wildlife immunogenetics: their relatedness to laboratory species means that well-defined mouse or rat genome data can be used to identify and characterise orthologous genes in the non-model species and, importantly, the use of well-studied natural rodent populations to study immunogenetics will allow immune gene variation to be analysed in a true ecological and evolutionary context.

2.2 OUTLINE

In this chapter I examine the genetic diversity within a wide range of immunity genes in a well-studied natural population of the field vole, *Microtus agrestis*. Single nucleotide polymorphisms within coding regions of these genes are described and characterised, and DNA sequence diversity estimated. Several molecular evolution analyses are performed on the sequence data to test for any departures from neutrality, which may be a result of natural selection acting on these genes.

2.3 MATERIALS AND METHODS

2.3.1 Candidate gene selection

The primary genes chosen for investigation were those that encode cytokines, signalling molecules that play a central role in the regulation of the immune response, while three non-cytokine genes were also chosen for study. The chosen genes were selected because of their known importance in the immune response and/or because previous research had implicated variation within these genes as being associated with infectious disease resistance (Table 2.1).

2.3.2 RNA extraction and cDNA synthesis

Vole spleens were obtained for RNA extraction as part of a separate study involving biannual monitoring and killing of animals, in March 2007. Spleens were dissected from twelve fresh, humanely killed voles and placed immediately in 10 mL RNA*later* (QIAGEN, Crawley, UK), incubated overnight at 4ºC and then stored at -80ºC until needed. Spleens from laboratory Wistar strain brown rat (*Rattus norvegicus*) (Charles River Laboratories, Margate, UK) from our laboratory were used to obtain RNA for subsequent positive controls. Total RNA was isolated from the spleen tissue using the TRIZOL Reagent method (Invitrogen, Paisley, UK) as per the manufacturer's instructions. The extracted RNA was quantified using an ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmingon, DE, USA) to measure absorbance at 260nm. Integrity of RNA was assessed via electrophoresis on a 1% agarose gel (Bioline, London, UK). Complementary DNA (cDNA) was synthesised from 1 μg total RNA per spleen using the SuperScript III Reverse Transcriptase method (Invitrogen), diluted 1:10 and stored at -20°C.

2.3.3 Primer design and amplification

Primers were designed to amplify gene fragments of up to ~1,000 bp in length from field vole cDNA. Because no public sequence data for vole immune genes are available, orthologous sequences from mouse, rat and human were identified using

Table 2.1 List of candidate field vole immune genes

Gene name	Gene symbol	Protein symbol	Summary
Cytokine			
Interferon gamma	*Ifng*	IFN-γ	The most important cytokine involved in the cell-mediated response. IFN-γ is secreted primarily by Th1 and NK cells in response to viral or bacterial infection, subsequently activating macrophages and generally up-regulating Th1-led cell-mediated immunity (Farrar and Schreiber 1993; Borish and Steinke 2003).
Interleukin 1, alpha	*Il1a*	IL -1α	A multifunctional cytokine produced by mononuclear phagocytes and endothelial cells (Sims and Smith 2010); involved in various aspects of the immune response including T-cell activation, macrophage stimulation and inflammatory responses (Borish and Steinke 2003).
Interleukin 1, beta	*Il1b*	IL-1β	An innate pro-inflammatory cytokine with identical biological functions to IL -1α, although IL-1β acts in a more systemic and less localized fashion (Sims and Smith 2010).
Interleukin 2	*Il2*	IL-2	Potent T-cell growth factor, produced by T-cells and acts in a paracrine way to induce proliferation. Also important in activation of B-cells, natural killer (NK) cells and macrophages (Borish and Steinke 2003).
Interleukin 4	*Il4*	IL-4	Important driver of naïve T-cells towards a Th2 phenotype and in resistance to helminths (Else *et al.* 1994; Borish and Steinke 2003).
Interleukin 5	*Il5*	IL-5	A Th2-specific cytokine involved in B-cell activation and eosinophil production (Clutterbuck *et al.* 1989; Borish and Steinke 2003).
Interleukin 10	*Il10*	IL-10	An anti-inflammatory cytokine secreted by regulatory T-cells, monocytes and B-cells among many others. Inhibits the production of many other cytokines involved in Th1, Th2 and pro-inflammatory responses (Borish and Steinke 2003; Steinke and Borish 2006).
Interleukin 12, alpha	*Il12a*	IL-12α	Codes for the smaller subunit (p35) of IL-12, a pleiotropic cytokine involved in pro-inflammatory responses including activation and induction of cytotoxicity in NK cells as well as activation and differentiation of both Th1 and Th2 cells (Romani *et al.* 1997; Borish and Steinke 2003). Acts in synergy with IL-18 (Yoshimoto et al. 1997).
Interleukin 12, beta	*Il12b*	IL-12β	Codes for the larger subunit (p40) of IL-12. Same biological activity as IL-12α. This subunit also forms part of the related cytokine IL-23, which has both similar and distinct biological roles to IL-12 (Oppmann *et al.* 2000; Kastelein *et al.* 2007).

Table 2.1 continued

Gene name	Gene symbol	Protein symbol	Summary
Interleukin 13	*Il13*	IL-13	Homologous to the Th2 cytokine IL-4 and shares many of its biological functions. Important in the allergic response and has been shown to have an important, IL-4- independent role in murine helminth infections (Grencis 2001).
Interleukin 18	*Il18*	IL-18	A pro-inflammatory cytokine that acts in synergy with IL-12 to induce IFN-γ production in T-cells; inhibits IL-4 dependent activity (Yoshimoto *et al.* 1997).
Transforming growth factor, beta 1	*Tgfb1*	TGF- β1	A member of the TGF- β family, and a potent anti-inflammatory and regulatory cytokine (Letterio and Roberts 1998).
Tumour necrosis factor	*Tnf*	TNF	A pleiotropic pro-inflammatory cytokine, secreted mainly by macrophages and induced most potently by bacterial lipopolysaccharide acting through TLR2 and TLR4 (Borish and Steinke 2003).
Non-cytokine			
Solute carrier family 11a member 1	*Slc11a1*	SLC11A1	Formerly known as *Nramp1*, this gene codes for an ion transporter in the endosomes of macrophages and has been associated with resistance to a range of infections in mice and humans (Vidal 1993; Malo *et al.* 1994; Blackwell *et al.* 2001).
Toll-like receptor 2	*Tlr2*	TLR2	The Toll-like receptors are a highly-conserved family of proteins which play a fundamental role in the innate immune response by recognizing pathogen-associated molecular patterns (PAMPs) expressed by infectious organisms, mediating the release of inflammatory cytokines and inducing an immune response. TLR2 interacts with TLR1 and TLR6 to recognize a variety of microbial ligands (Takeda *et al.* 2003).
Toll-like receptor 4	*Tlr4*	TLR4	Like TLR2, TLR4 is an important pattern-recognition receptor (PRR), which primarily recognizes LPS (Takeda *et al.* 2003).

the National Centre for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/) and aligned using ClustalW software (Thompson *et al*., 1994) on the European Bioinformatics Institute (EBI) website (http://www.ebi.ac.uk/clustalw/). ClustalW enables the identification of highly conserved areas within the gene sequence of different species. These conserved areas of sequence were used to manually design primers. The primers were ordered from Sigma-Aldrich (Haverhill, UK), ranged from 17-22 bp in length, contained from zero to three degenerate nucleotides and had melting temperatures (T_m) that ranged from 52.6ºC to 68.9ºC, as calculated by the Sigma website (Table 2.2). Intronic *Il1b* sequence data were obtained from genomic DNA (as described in Chapter 3) to allow a comparison between the properties of coding and non-coding sequence.

Gene fragments were amplified using the polymerase chain reaction (PCR). To optimise PCR conditions, 20µL reactions were performed over a gradient of annealing temperatures (T_a). The following 20µL reaction mix was used: 1 µL 1:10 dilution vole cDNA template, 1 µL forward primer (10µM) , 1 µL reverse primer (10µM), 10 µL Biomix Red (Bioline) (which contains an ultra-stable, non-proofreading *Taq* DNA polymerase, $MgCl_2$, ultra-pure dNTPs and an inert red dye for direct gel loading) and 7 µL AccuGENE molecular biology water (Cambrex, Verviers, Belgium). Negative, no-template controls replaced the vole cDNA with water and positive controls were run using rat cDNA, as *R. norvegicus* was one of the species whose sequence data was used to design primers.

Approximate PCR annealing temperatures for each primer pair were assumed to be 5 ºC lower than the mean of the two calculated melting temperatures; to find the optimum T_a, PCR reactions were run over a gradient from 6 ºC below to 4 ºC above the approximate T_a. PCR conditions were as follows: 95 ºC for five minutes, followed by 35 cycles of 95 ºC for 30 seconds, various T_a for 30 seconds, and 72 ºC for one minute, with a final extension time of 72 ºC for seven minutes.

Table 2.2 Primers and PCR conditions used to amplify gene products from cDNA. Nucleotide key: A = adenine; C = cytosine; G = guanine; T = thymine; M = A or C; R =A or G; W = A or T; Y = C or T; K = G or T.

Gene	Primers	T_m (ºC)	PCR T_a (ºC)	Product size (bp)
Ifng	For: GAG CCA RAT TRT CTC TTT CTA C	55.4	50.8	250
	Rev: GAC TCC TTT TCC GCT TCC	61.3		
Il1b	For: ATT GTR GCT KTG GAG AAG CTG	63.1	53.2	470
	Rev: CTT GWG AGG TGC TGA TGT ACC	62.2		
	For2: CAA GTG TCT GAA GCA GCY ATG	58.8	48.0	370
	Rev2: CTG ACG AAT GGG AAC ATC	53.7		
Il2	For: ARC AGY GCA CCY ACT TCA AG	56.8	55.4	390
	Rev: TGY TGA GAT GAT GCT TTG AC	58.8		
Il5	For: CAK TGG TGA AAG AGA CCT TG	59.7	53	290
	Rev: AAC TCT TGC AGG TAR TCT AGG	56.7		
Il10	For: TGC CAA GCC TTR TCK GAR ATG	57.9	54.3	250
	Rev: GGT TGA TGA AGA TGT CAA AYT	55.6		
Il12b	For: AGA TGC TGG CCA RTA CAC C	62.5	58.6	540
	Rev: AGG GAG AAG TAG GAA TGK GGA G	62.3		
Il18	For: TGG AAT CAG ACM ACT TTG GC	59.3	59.2	440
	Rev: GAT TTA TCC CCA TTT TCA TCC	60.7		
Slc11a1	For: ATC CTC CTC TGG CTG ACC	61.9	58.2	1060
	Rev: GGA GGC TGG GCA GGT AG	63.5		
Tgfb1	For: CAT CGA RGC CAT CCG NGG	68.9	57.9	950
	Rev: CGT AGT ACA CGA TGG GCA G	61.2		
Tlr2	For: GYG AAA ATY TGA TGG TTG AAG	58.5	50.0	1020
	Rev: AGA AGT CCA GTT CRT ACT TGC	56.7		
Tlr4	For: AAA TGG CTG GCA ATT CTT TC	62.8	58.0	1000
	Rev: AGT CKT CTC CAG AAG ATG TGC	60.0		
Tnf	For: ACC ATG AGC ACA GAA AGC AT	62.3	55.4	600
	Rev: CTT CTC CAG CTG GAA GAC T	52.6		

PCR products were electrophoresed on a 1% agarose gel (Bioline) to determine whether the primers had amplified single products of the expected size and which T_a was the optimum. Table 2.2 lists the optimum PCR annealing temperatures and amplified product sizes for the various genes. The genes *Il1a*, *Il4*, *Il12a* and *Il13* failed to amplify after several attempts to redesign primers and optimise PCR conditions and are therefore excluded from all further analyses.

2.3.4 Sequencing and SNP identification

Once the optimum T_a had been established, 50 µL PCR reactions, consisting of 3 µL 1:10 dilution vole cDNA template, 2.5 µL forward primer (10 µM), 2.5 µL reverse primer (10 µM), 25 µL Biomix Red (Bioline) and 17 µL molecular biology water (Cambrex), were performed to amplify sufficient product for sequencing, from 12 individual voles. Amplified products were run on a gel to confirm size, purified using MinElute PCR Purification Kits (QIAGEN), and quantified using an ND-1000 Spectrophotometer (NanoDrop Technologies). Samples were sent with appropriate primers to Eurofins MWG Operon (London, UK) for direct forward and reverse sequencing. As the *Taq* polymerase used in the PCR reactions was not a proof-reading enzyme, forward and reverse sequencing was performed on PCR products from separate reactions, to confirm that mutations were true SNPs and not artefacts from PCR errors. To confirm that the correct vole gene fragment had been amplified, a BLAST (Basic Local Alignment and Search Tool) search (Altschul *et al.* 1990) was performed on the obtained sequences to identify orthologous sequences in other species. The BLAST search also returns a score of sequence homology between species, measured as the percentage of nucleotide sites that are identical in both species.

Single nucleotide polymorphisms (SNPs) were identified manually from the sequence files obtained from MWG. Individuals homozygous for a SNP at a particular locus could be identified via the use of ClustalW (Thompson *et al.* 1994) and BioEdit (Hall 1999) software by aligning and comparing their sequence with other individuals, which highlighted any nucleotide differences. Heterozygous individuals were identified by manually checking each sequence trace using Chromas Lite version 2.01 (http://www.technelysium.com.au/chromas_lite.html), where SNPs could be readily detected as heterozygous peaks on the forward and reverse traces.

Each SNP was designated as non-synonymous (resulting in an amino acid change) or synonymous (no change in amino acid residue at that site) by aligning and comparing the derived vole cDNA sequences with known, orthologous mouse cDNA sequences retrieved from GenBank, assuming that no frame-shift mutations have occurred since the

two species diverged. Predictions to estimate the functional impact of each of the non-synonymous substitutions were conducted using PolyPhen (http://genetics.bwh.harvard.edu/pph/; Ramensky *et al.* 2002) and SIFT (http://sift.jcvi.org/; Kumar *et al.* 2009). Each program predicts the possible phenotypic impact of an amino acid substitution on the structure and function of the translated protein. SIFT uses data from multiple alignments to predict deleterious mutations, under the assumption that nucleotide sites important in protein function should be highly conserved between species, whereas relatively unimportant positions will be more diverse. PolyPhen also uses multiple alignments, but further accounts for the physical characteristics of the protein in its predictions, including information on the protein region in which the substitution occurs and the biochemical properties and compatibility of the replacement amino acid.

2.3.5 Sequence analysis

Analyses of sequence data were performed in DnaSP v5 (Librado and Rozas 2009). Nucleotide diversity was estimated by π, the average number of pairwise nucleotide differences between paired sequences (Nei 1987), and Watterson's θ_W (Watterson 1975), an estimator of the neutral population parameter θ (see below) based on the total number of segregating sites. Haplotype diversity (H_d), defined as the probability that two randomly chosen haplotypes are different in the sample, was estimated as

$$H_d = \frac{n}{n-1}\left(1 - \sum_{i=1}^{k} p_i^2\right),$$

where n is the number of gene copies in the sample, k is the number of haplotypes and p_i is the sample frequency of the i^{th} haplotype (Nei 1987). Several statistical tests of neutrality were conducted to identify any genes that departed from models of neutral evolution. Under the infinite sites model, which is typically used to model DNA sequence evolution and assumes that the number of nucleotide sites is large enough that each new mutation occurs at a site that has not previously mutated (Kimura 1969), expected nucleotide variation for a neutrally evolving diploid population is $\theta = 4N_e\mu$, where N_e is the effective population size and μ is the mutation rate. Tajima's (1989) D

reflects the difference between two estimators of θ, π and θ_W, the value of which is expected to be close to zero under neutrality. Significantly positive or negative values indicate a skew in the allele spectrum and may be due to selective or demographic effects.

The other neutrality tests required outgroup sequences; for each of these tests and for each locus two tests were performed, one using mouse as the outgroup sequence (as *Mus musculus* is the most closely related species to field vole that has had its genome fully sequenced) and one using the most closely related species for which sequence data were available for that particular gene. The mouse sequences therefore provided consistency of testing across all loci but at the expense of increased chance of multiple substitutions due to the relatively large divergence time between field voles and mice (~24.7 mya, Steppan *et al.* 2004), whereas the use of the most closely-related available species decreased the probability of multiple substitutions biasing results but with a loss of consistency across loci due to the variation in available sequence. The following species were used as the 'closest-relative' outgroup: bank vole, *Myodes glareolus*, (divergence time from field vole approximately 6.5 mya) amplified using field vole primers, was used as the outgroup sequence for *Il18, Tlr4* and *Tnf*; golden hamster, *Mesocricetus auratus* (divergence time ~18.5 mya), for *Il2* and *IL12b*; Chinese hamster, *Cricetulus griseus* (~18.5 mya), for *Tlr2*; hispid cotton rat, *Sigmodon hispidus* (~19.2 mya), for *Tgfb1*; while *Mus musculus* was the most closely-related species with sequence available for the gene *Slc11a1*. Using these outgroups, I then computed the following metrics: (1) Fu and Li's (1993) *D* and *F* tests, which are also based on comparison between two estimators of θ; the *D* statistic is based on the differences between η_e, the total number of mutations in external branches of the genealogy (recent mutations, inferred using an outgroup), and η, the total number of mutations, while the *F* statistic is the normalised difference between η_e and π. An excess of recent compared to older mutations (which occur on the 'internal' branches of the genealogy) may indicate negative selection (as deleterious alleles are present in low frequencies) or positive selection (as if an advantageous allele has recently been fixed in a population then the majority of the mutations in the population are expected to be young) (Fu and Li 1993). A deficiency of mutations in the younger, external branches and an increased number of older mutations may indicate balancing

selection. Because both alleles segregating at the Il1b 253 A/G SNP (Table 2.5) are different to the corresponding site in all outgroup sequences it was impossible to ascertain which field vole allele was the most recent; therefore, Fu and Li's tests were performed twice at this locus, once under the assumption that the *A* allele was most recent and once assuming it was the *G*. (2) Fay and Wu's (2000) *H*-test, a one-sided test for detecting positive selection on the basis of the relative excess of high-frequency-derived alleles expected immediately after a selective sweep. An excess of high-frequency derived alleles is a hallmark of genetic hitchhiking and positive selection (Fay and Wu 2000; Biswas and Akey 2006), and their test is based on the difference between π and θ_H, an estimator weighted by the homozygosity of the derived variants. (3) The McDonald-Kreitman (MK) test (McDonald and Kreitman 1991) which is based on the relative contribution of synonymous and non-synonymous substitutions within and between species.

In order to assess the significance of any departures from neutrality, empirical distributions of Tajima's, Fu and Li's and Fay and Wu's statistics were generated using neutral coalescent simulations (Hudson 1990) within DnaSP, based on the observed number of segregating sites and the simplifying assumption of no recombination, for 16,000 replicates. As each simulation results in a very slightly different *p*-value, an average was taken over 10 simulations. DnaSP returns a one sided *p*-value, reported as the proportion of coalescent simulations that had a test statistic equal to or lower than that observed. A one-sided test such as this would imply that a lower test statistic was suspected prior to the test; however, no such prior expectations are appropriate for these analyses. Therefore, for a two-tailed test, I took as significant any values where the observed test statistic was below the 2.5^{th} percentile or above the 97.5^{th} percentile of values obtained via the neutral coalescent simulations.

2.4 RESULTS

Twelve field vole immune gene fragments were successfully amplified and sequenced from between 8 - 12 individuals (16-24 haplotypes). In total, 6,629 bp of good quality

coding sequence data were obtained, yielding 26 SNPs; one per 255 bp of coding sequence (Table 2.3). A 547 bp fragment of non-coding *Il1b* intronic DNA was also sequenced for 12 voles and analysed for SNPs. Figures 2.1 and 2.2 show examples of an alignment and sequence trace, respectively, used to identify SNPs in the *Il1b* gene. Consensus sequences for the twelve coding regions were submitted to GenBank under the accession numbers HM245332 to HM245343 (Table 2.3).

The vole immune sequences covered between 36 – 86% of the total coding sequence of each gene - an average of 59% coverage - as estimated from the length of orthologous mouse coding sequences (Table 2.4). Estimates of nucleotide homology with other species for each field vole gene can be found in Table 2.4. The most highly conserved gene was *Il10*, with a 99% nucleotide homology to the corresponding bank vole (*Myodes glareolus*) sequence and an average of 94% homology over all compared sequences.

Table 2.3 Summary of sequenced field vole immune genes and GenBank accession numbers for consensus sequences, which include polymorphic sites, defined using IUPAC ambiguity codes

Gene	2*n*[a]	Length[b]	SNPs	Bp per SNP	GenBank accession no.
Ifng	24	220	0	na	HM245332
Il1b	24	695	3	232	HM245333
Il2	20	349	2	175	HM245334
Il5	20	242	0	na	HM245335
Il10	24	220	0	na	HM245336
Il12b	18	520	4	130	HM245337
Il18	24	410	1	410	HM245338
Slc11a1	16	592	2	296	HM245339
Tgfb1	16	937	1	937	HM245340
Tlr2	24	937	8	117	HM245341
Tlr4	18	926	3	309	HM245342
Tnf	22	581	2	291	HM245343
	Total	**6629**	**26**	**255**	
Il1b intron	24	547	8	68	

[a] Number of haplotypes sequenced
[b] Length sequenced for each locus

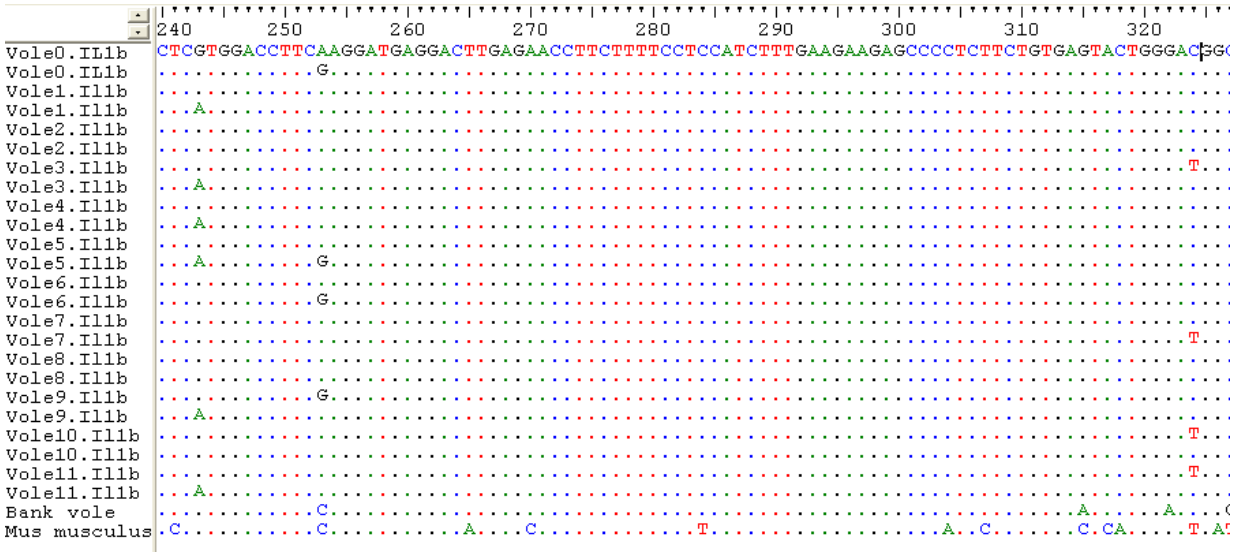

Figure 2.1 Section of *Il1b* alignment showing the three polymorphic sites in the field vole, and nucleotide differences in bank vole and mouse. Dots indicate nucleotides identical to the top-most field vole sequence.

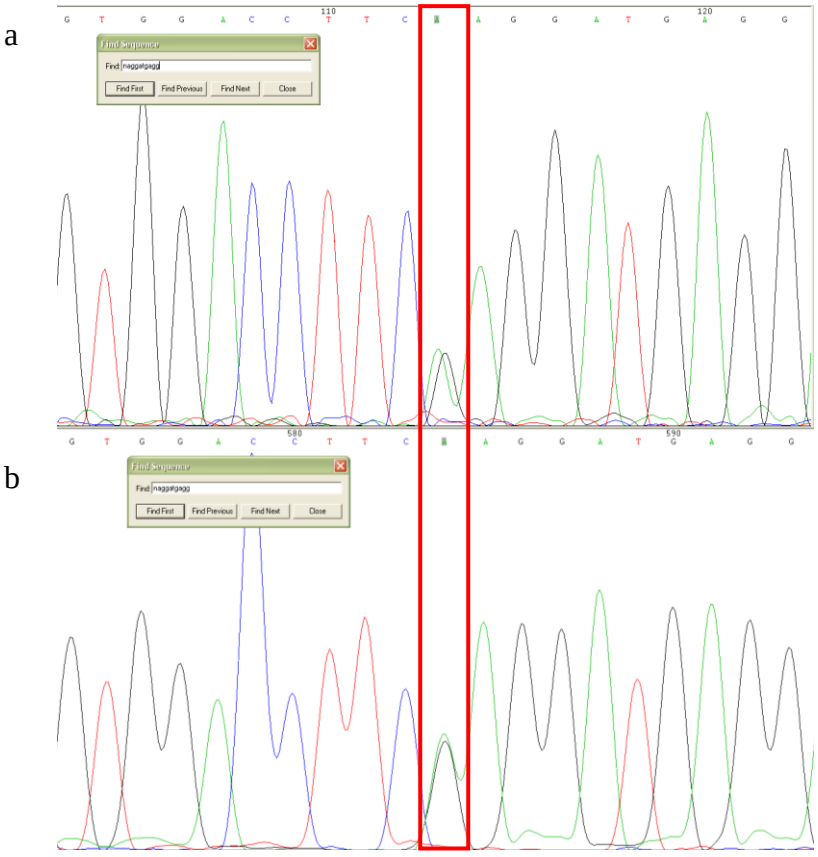

Figure 2.2 Forward (a) and reverse-complemented reverse trace (b) of an individual heterozygous at the highlighted site.

Table 2.4 Nucleotide homology between field vole and related species, ordered by time since species divergence. Estimates of divergence dates are (1) from Steppan *et al*. (2004) and (2) youngest possible date of divergence, from Benton and Donoghue (2007). Blanks, no available sequence data.

			Nucleotide homology %								
Gene	**Mouse CDS length** [a]	**Coverage (%)** [b]	*Myodes glareolus*	*Mesocricetus auratus*	*Peromyscus maniculatus*	*Sigmodon hispidus*	*Rattus norvegicus*	*Mus musculus*	*Meriones unguiculatus*	*Homo sapiens*	**Average**
Ifng	468	47.0	94	80	79	80	73	70	76	69	79
Il1b	810	85.8	95	89	-	88	85	85	87	75	86
Il2	510	68.4	-	89	87	86	85	82	83	77	84
Il5	402	60.2	-	-	92	89	87	89	89	80	88
Il10	537	41.0	99	95	95	92	94	93	95	88	94
Il12b	1008	51.6	-	89	-	88	83	81	78	76	83
Il18	579	70.8	80	-	-	80	98	91	84	75	85
Slc11a1	1647	35.9	-	-	-	-	89	89	-	85	88
Tgfb1	1173	79.9	-	89	93	92	91	90	92	84	91
Tlr2	2355	39.8	93	-	-	88	87	86	-	77	86
Tlr4	2508	36.9	95	-	-	-	87	87	-	79	87
Tnf	708	82.1	97	91	93	91	88	90	85	82	90
	Average	58.3	**93**	**89**	**90**	**87**	**87**	**86**	**85**	**79**	
		Divergence date (mya)	6.5[1]	18.5[1]	19.2[1]	19.2[1]	24.7[1]	24.7[1]	24.7[1]	61.7[2]	

[a] Complete coding sequence length of orthologous mouse gene (bp)

[b] Percentage of mouse coding sequence covered by aligned vole sequence

Ifng demonstrated the highest level of divergence between species at only 94% similarity with bank vole and an average of 79%. The accumulation of SNPs between species is expected to be roughly linearly correlated to divergence time (Sacks and Louie 2008) and although sequence data were not available for every gene in every species here, a strong negative correlation was indeed found between nucleotide homology and estimated time since species divergence (Steppan *et al.* 2004; Benton and Donoghue 2007) (r = -0.942, p <0.001).

2.4.1 Single nucleotide polymorphisms

Three gene amplicons – *Ifng*, *Il5* and *Il10* - did not contain any polymorphic sites. Of the nine genes that did contain at least one substitution, frequency of SNPs differed, ranging from 1 - 8 per gene and from one SNP per 117 – 937 bp of sequence (Table 2.3). The non-coding *Il1b* intronic sequence contained the highest level of polymorphism, with one SNP occurring per 68 bp of sequence.

A summary of all of the discovered SNPs can be found in Table 2.5; SNP positions are given relative to mouse reference sequences. Ten of the 26 SNPs resulted in an amino acid change in the translated protein. Of these ten non-synonymous substitutions only one, the Il12b 704 C/T SNP, was predicted to have a major effect on the translated protein; PolyPhen and SIFT predicted this mutation to be 'probably damaging' and to 'affect protein function', respectively.

2.4.2 Sequence diversity and tests for natural selection

The genetic diversity estimates and tests of neutrality (Table 2.6) illustrate the variation in levels of diversity within field vole immune genes. The highest levels of diversity were seen in the non-coding, intronic *Il1b* sequence, with high haplotype diversity (H_d) and estimates of nucleotide diversity (π) double that of the most polymorphic genetic coding region, *Tlr2*. Three other genes, *Il1b*, *Il18* and *Tnf* also demonstrated high haplotype diversities. Nucleotide diversity varied substantially between genes, with the lowest levels found in *Tgfb1* and *Tlr4* and the highest in *Il2*, *Tlr2* and *Tnf* (Table 2.6).

Table 2.5 Summary of SNPs found within a subset of field vole immune genes

Gene	SNP[a]	cDNA site[b]	gDNA site[c]	AA site[d]	Codon (amino acid) change	Syn/Nonsyn[e]
Il1b	Il1b 243 G/A	243	3109	81	TCG (Ser) → TCA (Ser)	Syn
	Il1b 253 A/G	253	3119	85	AAG (Lys) → GAG (Glu)	Nonsyn
	Il1b 324 C/T	324	3736	108	GAT (Asp) → GAC (Asp)	Syn
Il2	Il2 381 A/T	381	2949	127	ACA (Thr) → ACT (Thr)	Syn
	Il2 408 C/G	408	4748	140	AAC (Asn) → AAG (Lys)	Nonsyn
Il12b	Il12b 278 G/C	278	7935	93	GGA (Gly) → GCA (Ala)	Nonsyn
	Il12b 457 A/G	457	8500	153	ATA (Ile) → GTA (Val)	Nonsyn
	Il12b 465 T/C	465	8508	155	AGT (Ser) → AGC (Ser)	Syn
	Il12b 704 C/T	704	10901	235	CCG (Pro) → CTG (Leu)	Nonsyn
Il18	Il18 198 T/C	198	12503	66	GAT (Asp) → GAC (Asp)	Syn
Slc11a1	Slc11a1 537 C/G	537	5014	179	TTC (Phe) → TTG (Leu)	Nonsyn
	Slc11a1 714 G/A	714	5697	238	CTG (Leu) → CTA (Leu)	Syn
Tgfb1	Tgfb1 442 C/A	442	5462	148	CCA (Pro) → ACA (Thr)	Nonsyn
Tlr2	Tlr2 1383 G/A	1383	4376	461	GTG (Val) → GTA (Val)	Syn
	Tlr2 1614 C/A	1614	4607	538	TCC (Ser) → TCA (Ser)	Syn
	Tlr2 1648 G/A	1648	4641	550	GCA (Ala) → ACA (Thr)	Nonsyn
	Tlr2 1706 G/A	1706	4699	569	CGC (Arg) → CAC (His)	Nonsyn
	Tlr2 1821 C/T	1821	4814	607	GGC (Gly) → GGT (Gly)	Syn
	Tlr2 1852 C/T	1852	4845	618	CTG (Leu) → TTG (Leu)	Syn
	Tlr2 1905 G/A	1905	4898	635	CCG (Pro) → CCA (Pro)	Syn
	Tlr2 1944 T/C	1944	4937	648	AGT (Ser) → AGC (Ser)	Syn
Tlr4	Tlr4 1663 A/G	1663	12824	555	AAG (Lys) → GAG (Glu)	Nonsyn
	Tlr4 1848 G/T	1848	13009	616	Leu (CTG) → Leu (CTT)	Syn
	Tlr4 2037 C/A	2037	13198	679	Ser (TCC) → Ser (TCA)	Syn
Tnf	Tnf 138 A/G	138	294	46	Thr (ACA) → Thr (ACG)	Syn
	Tnf 210 T/C	210	882	70	Ile (ATT) → Ile (ATC)	Syn

[a] SNP designation includes the encompassing gene, SNP position relative to aligned mouse cDNA sequence and the resulting base change

[b] Position of SNP relative to aligned mouse cDNA sequence

[c] Position of SNP relative to aligned mouse genomic DNA sequence

[d] Position of SNP codon relative to aligned mouse amino acid sequence

[e] Denotes whether the SNP is synonymous or non-synonymous

Table 2.6. Sequence diversity and tests of neutrality. *Ifng*, *Il5* and *Il10* sequences are excluded as no polymorphisms were found within these genes. No significant results were obtained using the MK test.

Gene	N_h [a]	H_d [b]	π [c]	θ_W [d]	Tajima's *D*	Fu and Li's *D*	Fu and Li's *F*	Fay and Wu's *H*
Coding								
Il1b	5	0.73	14	11.6	0.525	**0.828*** **	1.002	na
Il2	2	0.43	24.5	15.8	1.126	**0.871 *** **	1.093	0.523
Il12b	5	0.53	17.6	21.8	-0.750	0.199	-0.069	-0.833
Il18	2	0.52	12.8	6.5	1.505	0.640	0.994	0.104
Slc11a1	2	0.36	18.4	15.9	0.415	0.915	0.921	0.527
Tgfb1	2	0.13	2.3	5.5	-1.162	-1.523	-1.657	0.117
Tlr2	9	0.87	23.6	29.7	-0.696	-0.475	-0.645	0.842
Tlr4	3	0.24	4.3	10.3	-1.697	-1.273	-1.626	-1.517
Tnf	4	**0.69 ***	22.5	12.8	**1.935 ***	**0.891 *** **	**1.353 ***	0.117
Average (SD)	3.7 (2.3)	0.50 (0.24)	15.1 (7.9)	14.4 (7.6)	-0.320 (1.02)	-0.253 (1.01)	0.002 (1.186)	0.102 (0.767)
Non-coding								
Il1b intron	11	**0.93 ****	46.2	40.1	0.499	0.509	0.602	-1.316

[a] Number of haplotypes
[b] Haplotype diversity
[c] Mean number of pairwise differences per site (x 10^4)
[d] Watterson's estimator (x 10^4). Significant values in bold. Stars indicate significance level:
* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; *p*-values generated by neutral coalescent simulations assuming no recombination, using DnaSP

Evidence for non-neutral evolution was seen for the genes *Il1b*, *Il2* and *Tnf* (Table 2.6) which all gave significant results for at least one of the tests of neutrality, and all in the same direction. Fu and Li's *D* was significantly positive for *Il1b*, indicating an excess of mutations in the older, internal branches of the genealogy. Fu and Li's tests utilise an outgroup sequence to infer which alleles are derived and which are ancestral. This was uncertain for *Il1b*; however the results did not differ between whether the *A* or *G* nucleotide was assumed to be the derived allele. However, results did differ for Fay and Wu's *H* test under the two assumptions and were therefore excluded. Fu and Li's *D* value was also significantly high for *Il2* ($p < 0.001$) while *Tnf* had consistently high values across Tajima's ($p = 0.012$) and Fu and Li's *D* ($p < 0.001$) and *F* ($p = 0.013$) tests, each indicative of an excess of internal, intermediate frequency alleles. There was no difference in the results of

the neutrality tests that utilized an outgroup sequence when *Mus* or the closest available relative was used as the outgroup sequence. Also of note were the test values for *Tlr4* which were consistently low and borderline non-significant ($0.05 < p < 0.06$) for Tajima's *D*, Fu and Li's *F* and Fay and Wu's *H* tests. The McDonald-Kreitman test did not identify any departures from neutrality.

2.5 DISCUSSION

This study is the first to describe the genetic diversity within a wide range of immune function genes in a natural population. Previous studies on wildlife immunogenetics have concentrated almost exclusively on genes of the MHC (Acevedo-Whitehouse and Cunningham 2006), with only a small number of such studies describing diversity in other immune genes. Here I sequenced partial coding regions from twelve non-MHC immunity genes and one intronic region of the field vole, an ecologically well-studied non-model species, and characterised a number of single nucleotide polymorphisms.

Genetic diversity was highest within the non-coding region and varied markedly between the coding sequences. Molecular evolution analyses provided evidence of non-neutral evolution at several of these genes, which may result from selective or demographic processes. This immunogenetic data provides a valuable resource for further studies on the molecular ecology and evolution of host-parasite interactions in the wild.

2.5.1 Genetic diversity

Levels of genetic diversity differed among the immune genes. As one would expect, the highest levels of diversity were seen in the intronic region of *Il1b*, reflecting the relaxation of selection at this non-coding locus. The lower levels of genetic diversity in the coding regions compared to the intron indicates that even the most diverse of the functional genes have been subject to some level of negative or background selection acting to prevent change and to conserve function. No

polymorphic sites were found within the three genes *Ifng*, *Il5* and *Il10*; this in part was no doubt due to the fact that it was possible to sequence only relatively short lengths for these genes (220 – 242 bp). However, the low diversity observed within these loci is likely to have also been shaped by selection; the fact that *Il10* in particular is so highly conserved between species suggests that negative selection has acted to prevent molecular changes at this locus. *Ifng* shows a more unusual pattern of diversity; low levels of within-species diversity at this locus have been seen in this study and in other species (Worley *et al.* 2006; Downing *et al.* 2009), which contrasts with the much higher number of between-species differences at this gene (Table 2.4). High between-species and low within-species diversity may be indicative of the action of positive selection, with selective sweeps driving divergence between species while reducing the level of diversity around the site of selection within a species (Maynard Smith and Haigh 1974). A recent paper by Levi-Acobas *et al*. (2009) has found evidence for several episodes of strong, positive Darwinian selection acting on the *Ifng* gene in the Glires lineage, a mammalian clade comprising of lagomorphs and rodents, including voles. They also examined genetic diversity within the *Mus* genus and found evidence for a relatively recent selective sweep affecting *Mus musculus domesticus*, as evidenced by a much lower level of diversity in a 50 Kb chromosomal segment encompassing the *Ifng* gene when compared to other *Mus musculus* subspecies. Although their study involved comparisons between laboratory rather than natural species, these findings support the idea of past episodes of positive selection as a possible mechanism for explaining the patterns of diversity seen within and between field voles and related species at the *Ifng* gene.

The highest level of coding-region genetic diversity was seen in *Tlr2*. A far higher proportion of the observed nucleotide variation at this gene was at synonymous (π_s = 8.91 x 10^{-3}) compared to non-synonymous sites (π_{ns} = 3.5 x 10^{-4}). As synonymous substitutions are generally assumed to be invisible to natural selection (but see Akashi 1995) and because of the lack of significance for any of the tests of neutrality, this relatively high level of diversity conforms to neutral expectations. Unusually high levels of haplotypic diversity (Table 2.6) were also seen in the

genes *Il1b* and *Tnf* which may be due to relaxed selective constraint, drift, or natural selection acting to maintain diversity at these regions.

2.5.2 Evidence for natural selection

Evidence for non-neutral evolution was observed for the genes *Il1b*, *Il2* and *Tnf*, through the use of population genetic tests of neutrality. *Il1b* had a high value for Fu and Li's *D*, which is indicative of an excess of internal mutations and a sign of balancing selection. Interestingly, the two alleles of the nonsynonymous Il1b 253 A/G SNP were both different to any of the outgroup sequences. This suggests a relatively recent, field vole-specific mutation at this site, followed by another. This site is conserved in all closely-related species; mouse, rat, Mongolian gerbil, hispid cotton rat, golden hamster and bank vole sequences all have a cytosine at this site. It would seem that the chances of two derived mutations being retained at this nucleotide site in the field vole by chance are slim; one explanation may therefore be that both field vole alleles offer a fitness advantage over the ancestral cytosine and that some form of balancing selection maintained both in the population. Later chapters will analyse any link between genotype and phenotype at this locus to ascertain whether there are any phenotypic consequences of the polymorphism found within the *Il1b* gene. Subsequent chapters will also examine any phenotypic association with the Il12b 704 C/T SNP which was predicted to be functionally important.

Fu and Li's *D* was also significantly positive for *Il2*, while *Tnf* had consistently high values across several tests of neutrality, supporting the action of balancing selection. No departures from neutrality were detected by the MK-test, which is primarily a test for directional rather than balancing selection. The MK test also measures long-term evolution between species rather than contemporary selection acting within a species; although specific examples of long-term balancing selection exist, such as that maintaining high allelic diversity at the MHC, it is probably uncommon (Charlesworth 2006). *Tlr4* represents the best candidate for a gene under directional selection; this gene was highly negative and close to significance ($0.05 < p < 0.06$) across several tests of neutrality, which is suggestive of positive

selection; interestingly, the derived allele at the Tlr4 1663 A/G SNP is the most common, which may indicate that positive selection has acted to drive the increase in frequency of this more recent allele.

It is possible to fully distinguish selection from demographic effects only by comparing multiple sites in the genome (Nielsen 2005). Because multiple genetic loci were tested for neutrality here, one may be more confident that the local signs of increased diversity in these genes are due to selection acting to maintain diversity, rather than demography; if demographic events had shaped the patterns of diversity seen within a population, one would expect the various neutrality test statistics to head in the same direction for all genes, reflecting the genome-wide effect that demography has on genetic variation. This was not the case in this study, where approximately even numbers of positive and negative values were seen for each of the neutrality tests.

Another possible explanation for the patterns of diversity seen in *Il1b*, *Il2* and *Tnf* is that they have arisen through ascertainment bias. Many studies of SNP data, including the present study, suffer from some degree of ascertainment bias in that a relatively small sample of individuals are used to discover SNPs before genotyping a much larger sample; this means that rare SNPs are less likely to be discovered (Nielsen and Signorovitch 2003; Morin *et al.* 2004). SNP discovery in small samples does increase the likelihood of the data being unrepresentative of the allele frequency spectrum of the population and having a relative excess of intermediate frequency alleles (Nielsen and Signorovitch 2003). In this study some degree of bias was necessary, as SNPs segregating at very low frequencies would have little power for detecting associations with parasitic disease or immune function (later chapters). However, extensive ascertainment bias would affect every gene studied and would therefore be likely to lead to an excess number of intermediate alleles over all loci; the fact that this was not the case in this study suggests that bias is not the reason for the excess number of intermediate alleles in the three genes. In addition, sample size in this study was large enough to potentially detect SNPs

segregating at relatively rare frequencies of 4–6% (and in fact SNPs segregating at even lower frequencies were discovered; Chapter 3).

2.6 CONCLUSION

This study has aimed to broaden research on wildlife immunogenetics away from concentrating solely on the MHC, by examining genetic diversity within a range of other functional immune genes in a well-studied natural population of field voles.

Genetic diversity within field vole immune genes has been shown to be widespread and variable. The functional relevance of this variation is alluded to by the evidence for balancing selection acting to maintain the variation observed at several of these genes. Further support for balancing selection and for the functional consequences of this genetic variation will be assessed in subsequent chapters, through the examination of population-level factors and genotype-phenotype associations. More broadly, this study demonstrates the potential of using genomic data of well-characterised laboratory species in order to investigate important ecological and evolutionary questions in related non-model organisms.

Chapter 3

Population genetics and natural selection at immune gene loci

3.1 INTRODUCTION

Understanding the mechanisms responsible for the generation and maintenance of genetic diversity is a central objective of evolutionary biology. Levels of genetic variation that deviate from those expected under neutrality can result from both population demographic and selective events and disentangling these processes is an important and challenging task for evolutionary biologists (Beaumont and Balding 2004). In an attempt to do so, the field of population genomics has arisen through the combination of population genetics with genomic concepts and methods (Luikart *et al.* 2003). Population genomic approaches rely on the simultaneous study of many loci across the genome in order to separate locus-specific effects, which imply the action of natural selection, from genome-wide effects such as inbreeding or genetic drift which should affect all parts of the genome in the same way (Black *et al.* 2001). Identifying the genes in natural populations that are the targets of selection has particular implications for conservation genetics, in which maintenance of genetic diversity is of central importance (Sommer 2005; Smith *et al.* 2009), but also improves our understanding of adaptive evolutionary processes as a whole.

A large body of evidence suggests that parasites are a major selective force in maintaining host genetic diversity in both human (reviewed in Barreiro and Quintana-Murci 2010) and wild populations (reviewed in Sommer 2005). Genes of the immune system may therefore be particularly likely to experience strong selective pressures, as these are the genes that interact with the myriad of pathogens that constantly challenge free-living organisms. The vast majority of studies addressing adaptive evolution within immune genes in natural populations have concentrated on the MHC, where parasite-driven natural selection has acted to maintain extremely high levels of genetic diversity (Piertney and Oliver 2006).

There is however a need to examine a wider range of immune-function genes, which will allow for a much broader understanding of the effect natural selection has on immunity, genetic diversity and fitness in natural populations (Acevedo-Whitehouse and Cunningham 2006). This will allow for an improved understanding of infectious disease dynamics in wild species and may in turn inform decisions on conservation and epidemiology. Thus far, however, very few studies have attempted to characterise how and why natural selection acts on non-MHC immune genes in wild species (Acevedo-Whitehouse and Cunningham 2006; Worley *et al.* 2006; Jensen *et al.* 2008; Tonteri *et al.* 2010).

The previous chapter described the DNA sequence variation of a subset of field vole immune genes and used these data to test for putative evidence of natural selection. Examining variation at the DNA sequence level can provide insights into the long-term processes that, over evolutionary timescales, have shaped within-species diversity and driven the divergence between species. For a more complete understanding of immunogenetic diversity and the processes which shape it, it is important to also investigate genetic variation at the population level. Natural selection operating in contemporary populations is expected to have an effect on allele frequencies within that population. Therefore, examination of the frequencies of alleles within and between populations enables us to better understand the processes that have shaped the genetic diversity within a species or individual population, over relatively recent timescales.

3.2 OUTLINE

This chapter examines the population-level genetic diversity of a range of non-MHC immune genes in the field vole. I present the results of genotyping several hundred individuals in a number of subpopulations located in neighbouring Kielder, Kershope and Redesdale Forests. I use these data to investigate allele frequencies within each population, population structure, and Hardy-Weinberg and linkage equilibria; linkage disequilibrium within individual genes is a prerequisite for any phenotypic associations caused by linked mutations to be detected and for

haplotype inference to be robust. These data are also used to provide further evidence for natural selection acting upon several of these immune genes, and thus shaping the patterns of genetic diversity within a natural field vole population.

3.3 MATERIALS AND METHODS

The primary practical aim of this chapter was the genotyping of SNPs discovered as described in the previous chapter, in several hundred field voles from several populations in and around Kielder Forest (see below). This genotyping data was then used as a basis for several standard population genetic analyses, including tests for Hardy-Weinberg and linkage equilibria, inbreeding and population subdivision with particular emphasis was placed on acquiring evidence for natural selection.

3.3.1 Genomic DNA sequencing

As SNP typing was to be performed from gDNA rather than cDNA, at least 50bp of gDNA sequence either side of each SNP was required to design primers for genotyping. Genomic DNA was extracted from spleen tissue from two voles used for SNP discovery (Chapter 2). Firstly an extraction buffer solution was prepared consisting of 0.1M NaCl, 20mM Tris at *p*H 8.0, 25mM EDTA disodium salt at *p*H 8.0 and 0.5% SDS in ddH_2O. This was filtered through a 0.45 μm Minsart syringe end filter (Sartorius, Epsom, UK) and autoclaved at 121ºC for 15.5 minutes. Tissue samples were placed in 0.3 mL extraction buffer and crushed. Extraction buffer was then added up to 0.5 mL, followed by addition of 50 μl of 10 mg mL^{-1} proteinase-K. The tube was inverted to mix, and incubated at 55ºC overnight or until the tissue was completely digested. After digestion, 0.8 mL of PCI (phenol:chloroform:IAA) (Sigma) was added, the solution vortexed and centrifuged for 5 minutes at 12,000 rpm. The resulting aqueous phase was aspirated into a fresh tube, 50 μl 3M NaAC *p*H 5.3 was added and the solution was mixed. 100% ethanol was then added up to 1.5 mL and the tube stored at -20ºC for approximately 30 minutes to allow the DNA to precipitate. The sample was then centrifuged at 12,000 rpm for 10 minutes and the supernatant removed by aspiration. The

resulting DNA pellet was washed in 1 mL 75% ethanol and stored at 4ºC for 30 minutes. The sample was spun at 12,000 rpm for 4 minutes, the supernatant was again aspirated and the pellet allowed to air dry in a fume cupboard. Finally the pellet was resuspended in TE solution and quantified using a ND-1000 Spectrophotometer (NanoDrop Technologies).

Primer design and PCR

Primers were designed by eye from the vole cDNA sequences obtained earlier. As genomic DNA contains non-coding introns as well as coding regions, primers were designed to amplify products from 100bp to 2Kb as necessary, depending on the length of unknown intronic sequence between the primer sites in the known coding regions, in order to obtain at least 50bp sequence either side of each SNP. Mouse genomic DNA sequence was used as a template to estimate the size of amplified PCR products in cases where introns occurred between primer sites. PCR conditions were optimised as in Chapter 2; Table 3.1 lists the PCR conditions and primers used to amplify around each SNP. For each reaction, 1 µL 1:10 dilution field vole gDNA was used as template, except for the Slc11a1 537 C/G SNP, which used 3 µL. PCR extension time was 1 minute for all reactions except for *Il2*, which had a two minute extension step.

3.3.2 Genomic DNA extraction for SNP typing

Tissue samples were obtained for gDNA extraction and genotyping from one previous and two concurrent studies. Livers were extracted from voles killed as part of a cross-sectional study investigating field vole immunodynamics, between March 2008 and March 2009 (Jackson *et al.* in review; see Chapter 4 for more details on trapping regime). Tail bleeds were obtained from a parallel longitudinal study involving capture-mark-recapture methods within the same sample areas as the cross-sectional study, from March to October 2008 (Chapter 4). Finally, vole tails were obtained in March 2008, from a long-term cross-sectional study involving twice-yearly sampling and killing of field vole populations in Kielder and surrounding forests (refer to Telfer *et al*. (2007a) for trapping methods).

Table 3.1 Primers and PCR conditions used to amplify around each SNP from vole gDNA. Two SNPs, Il12b 278 G/C and Il18 198 T/C, did not amplify from genomic DNA.

Gene	SNP	Primers	T_m (ºC)	PCR T_a (ºC)	Product size (bp)
Il1b	Il1b 243 G/A	For: CTG TTG ATC TGA GCT GTC CAG	62.7	60	780
	Il1b 253 A/G	Rev: GCT GAC GAA TGG GAA CAT C	63.2		
	Il1b 324 C/T				
Il2	Il2 381 A/T	For: GCC ACA GAA CTA AAA CAT CTT C	59.3	55	2000
	Il2 408 C/G	Rev: GTT GAG ATG ATG CTT TGA CAG	59.3		
Il12b	Il12b 457 A/G	For: AAT TAC TCC GGA CGT TTC ACC TGC TG	72.1	65.2	180
	Il12b 465 T/C	Rev: ATC CTC CTG GCA TGC AAC TGT GTA CTT C	73.0		
	Il12b 704 C/T	For: CTC TGA TTC CCG AGC AGT AAC	62.7	55	880
		Rev: CTC CAC CTC AGA GTT CCT CAG	63.0		
Slc11a1	Slc11a1 537 C/G	For: GTC CTG ATC ACC ATC GTA G	57.9	55	460
		Rev: CGT AAC CGA AGG TCA GAG	58.3		
	Slc11a1 714 G/A	For: CTC TGA CCT TCG GTT ACG	58.4	55	1150
		Rev: GAT GGT GGC TTC AAT CAG	59.1		
Tgfb1	Tgfb1 442 C/A	For: AAC CAA AGA CAT CCC ACA C	59.9	55	120
		Rev: GTA GAG TTC CAC GTG TTG TTC	58.7		
Tlr2	Tlr2 1383 G/A	For: TGA CAT CAG CCG GAA CAG	63.8	55	820
	Tlr2 1648 G/A	Rev: ATG GAG TCG ATG ATG TTG TC	60		
	Tlr2 1706 G/A				
	Tlr2 1944 T/C				
Tlr4	Tlr4 1663 A/G	For: ACA CAC TCC ATG GAC TTC TG	60.2	55	760
	Tlr4 1848 G/T	Rev: TCT CCA GGA CAA TGA AGA TG	60.2		
	Tlr4 2037 C/A				
Tnf	Tnf 138 A/G	For: AAC TCC AGT CGC TGC CTG	64.4	62	900
	Tnf 210 T/C	Rev: GGC TAC AGG TTT GTC ACT TG	59.7		

Livers

Genomic DNA was extracted from the livers of field voles caught and killed between March 2008 and March 2009 as part of an ongoing cross-sectional study (Jackson *et al.* in review). Three hundred and seven individuals were caught from two sites within Kielder Forest, designated BLB (55.2457, -2.6108; *n* = 152) and SQC (55.2549, -2.6116; *n* = 155) (Figure 3.1). Genomic DNA was extracted from the liver tissue using the DNeasy Blood and Tissue Kit (QIAGEN).

Tail bleeds

As part of an ongoing longitudinal study monitoring field vole infections which was running parallel to the cross-sectional study above, blood samples were obtained from tail-bleeds of 352 animals from the two sites within Kielder Forest, BLB (*n* = 153) and SQC (*n* = 192). Blood samples were placed immediately in 500 µl RNA*later* (QIAGEN), kept cool and stored at -80ºC as soon as possible after returning from the field. Genomic DNA was extracted from the tail blood/RNA*later* samples using DNAzol BD Reagent (Invitrogen), a guanidine-detergent lysing solution designed for extracting DNA from whole blood. Specifically, the protocol used was that of Mackey *et al*. (1998), which is modified for extracting DNA from small volumes of blood (less than 20 µL). Briefly, 15 µL of the blood/RNA*later* was added to 200 µL DNAzol BD and 1 µL GenElute linear polyacrylamide (Sigma) in a 1.5 mL microcentrifuge tube and left for 5 minutes at room temperature. 80 µL isopropanol was added, mixed by shaking, and left for a further 5 minutes before being centrifuged at 5000 *g* for 5 minutes at 8ºC. The supernatant was decanted and the tubes stored inverted for 3 minutes. The DNA pellet was then washed twice with 750 µL 95% ethanol, any residual ethanol removed, and the pellet solubilised in 12.5 µL sterile water. The DNA was quantified using an ND-1000 Spectrophotometer (NanoDrop Technologies). As the amount of gDNA extracted from these small samples was too low to genotype a large number of SNPs, 1 µL of the extracted gDNA was used as a template for whole genome amplification using the illustra GenomiPhi V2 DNA amplification kit (GE Healthcare, Buckinghamshire, UK).

Tail snips

Field vole tails were obtained as part of a separate, long-term cross sectional study (see above) involving biannual (March and September) monitoring and killing of field voles, in March 2008. This study utilised a much wider area, including sites in surrounding forests as well as in Kielder itself. Twenty-seven grass-dominated clearcut sites (5-12ha) within three adjacent man-made spruce forests were used for trapping. There were twelve sites located in Kielder Forest (55° 13'N, 2° 33'W), ten sites within Kershope Forest (55° 06'N, 2° 45'W) and five sites within Redesdale Forest (55° 17'N, 2° 21'W) (Figure 3.1). In total, 155 field voles were caught; 50 from Kielder Forest, 48 from Kershope and 57 from Redesdale. Genomic DNA was extracted from the tails using the DNeasy blood and tissue kit (QIAGEN). A more wide-ranging sampling area was used to allow a comparison of the genetic structure among and between Kielder and more distant populations of field voles. A March, rather than September, sample was chosen for the genetic analyses as March is at the beginning of the breeding season and samples are therefore less likely to contain families of young voles which could lead to spurious inferences of genetic structure (S. Telfer, personal communication).

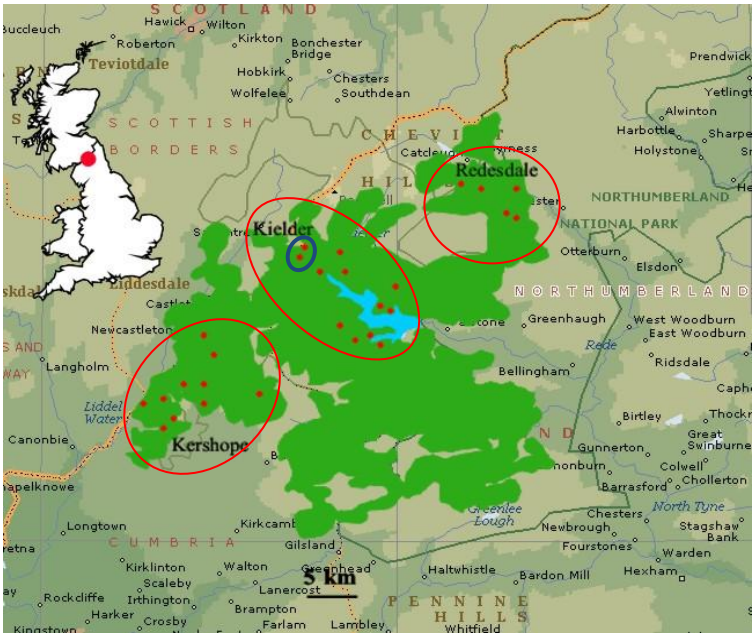

Figure 3.1 Sites used for trapping and acquisition of field vole tissue samples. Blue outline contains the within-Kielder sites, SQC (upper) and BLB (lower); red outlines represent the wider Kershope, Kielder and Redesdale Forest sites.

Genotyping was performed by KBioscience (Hoddesdon, UK; http://www.kbioscience.co.uk) utilising the KASPar SNP genotyping system, their own novel form of competitive allele-specific PCR (http://www.kbioscience.co.uk/chemistry/chemistry_Kasp_intro.html). In total, 817 individuals were genotyped at each of the SNP loci. Samples were provided in volumes of 100 µl and at a concentration of ~20 ng μl^{-1} as quantified by an ND-1000 Spectrophotometer (NanoDrop Technologies). This was a deliberately high concentration to allow for some inaccuracy of the spectrophotometer and also to enable further downstream dilutions if necessary. At least 50 bp genomic sequence around either side of each SNP was also provided. A number of samples were included as controls: twenty-one samples were provided in duplicate as positive controls to test for reliability of genotyping and several DNA samples from one individual of known genotype were provided on separate plates to check for reliability between plates. This individual was also used in triplicate for a positive control for the post-whole genome amplification genotyping. Several negative controls were provided; a pooled sample of all negative controls from whole genome amplification reactions (where the DNA template was replaced by water), water alone, and KBioscience also provide their own no-template controls. One sham 'SNP' was included as a check for miscalling rates – the *Il10* locus actually contained no SNPs; however, KBioscience were instructed to genotype a particular nucleotide site for an A/G SNP; in fact this nucleotide was exclusively an adenine. In total, 22 individuals were genotyped in duplicate for 19 SNPs (including the sham *Il10* SNP), giving a total number of 418 replicated genotypes.

Not all of the discovered SNPs (Chapter 2) were genotyped. Only four of the eight SNPs discovered in the *Tlr2* locus were genotyped, the two nonsynonymous SNPs and those at either end of the sequenced region. The Il12b 457 A/G, Il12b 465 T/C, Il18 198 T/C and Tnf 138 A/G SNPs could not be genotyped even after redesigning assays. In total 18 of the 26 discovered SNPS were able to be genotyped.

3.3.4 Hardy-Weinberg and linkage equilibria

GENEPOP 4.0 (Guo and Thompson 1992; Raymond and Rousset 1995; Rousset 2008) was used to calculate observed and expected heterozygosity and to test for

deviations from Hardy-Weinberg equilibrium using the Markov chain method to evaluate exact tests. Values of the inbreeding coefficient F_{IS} (Wright 1951), the mean reduction in heterozygosity of an individual due to non-random mating in a subpopulation, were calculated within Arlequin v. 3.11 (Excoffier *et al.* 2005) for each allele and over all alleles; the significance of F_{IS} values was then tested for using 10,000 permutations of the data.

The extent of linkage disequilibrium between SNPs was analysed using LinkDos (Garnier-Gere and Dillmann 1992). Genotypic linkage disequilibrium (Weir 1996) was determined as a measure of the association between genotypes rather than alleles. LinkDos uses composite frequencies and estimates the unbiased linkage disequilibrium coefficient between two alleles as

$$\Delta_{ij} = (N/N - 1)((T_{ij}/N) - 2p_ip_j),$$

where T_{ij} is the number of times that the alleles i and j appear in the same individual, N is the total number of genotypes, while p_i and p_j indicate the frequency of the i^{th} allele at one locus and the j^{th} allele at another locus, respectively (Black and Krafsur 1985). A correlation coefficient is determined,

$$R_{ij} = \Delta_{ij} / ((p_i(1 - p_i)+C_i)\ (p_j(1 - p_j)+C_j))^{1/2},$$

where the C term is a correction for departures from random mating; C_i is equal to the observed minus the expected frequencies of homozygotes for the i^{th} allele (Black and Krafsur 1985). A chi-squared statistic and corresponding significance level of R is calculated for each pair of alleles by

$$\chi^2 = NR_{ij}^2$$ (Weir 1979).

Because of the large number of tests performed, a sequential Bonferroni correction (Holm 1979; Rice 1989) was applied to the resulting p-values. Pairwise linkage disequilibrium measurements were calculated for every pair of SNPs under the

hypotheses that SNPs within the same gene would demonstrate high LD through linkage, and thus have greater power to pick up a phenotypic association if a causal mutation was up or downstream from the genotyped polymorphism, and that SNPs located in different genes would not demonstrate LD and that genetic loci were therefore independent. To ascertain whether significant LD observed between SNPs within different genes was due to epistatic selection or genetic drift between subpopulations, variance components of LD were calculated following the method of Ohta (1982). Ohta's theory is based on the hypothesis that if selection is operating and producing specific combinations of alleles then those combinations should appear consistently among subpopulations (Ohta 1982; Black and Krafsur 1985): she defines D_{IS}^2 as the expected variance of LD within a subpopulation, while D_{ST}^2 is the variance of the correlation of the genes of two loci of different gametes in a subpopulation, relative to the total population. D_{ST}^2 is therefore a predictor of the variance of LD expected to occur under random genetic drift. If subpopulations are differentiated, combinations of alleles would have been established independently among the subpopulations and therefore $D_{ST}^2 > D_{IS}^2$; conversely, uniform natural selection across all subpopulations will cause the frequency of allele combinations to converge so that $D_{ST}^2 < D_{IS}^2$ (Ohta 1982; Black and Krafsur 1985).

3.3.5 Population differentiation

The extent of genetic differentiation between subpopulations was assessed in two ways, both within the GENEPOP program; first, by estimating pairwise F_{ST} values following Weir & Cockerham (1984) and, second, testing for genotypic differentiation by examining the distribution of genotypes in the various populations. Here, the Markov chain method is used to evaluate exact *G* tests and produce unbiased *p*-value estimates under the null hypothesis of "genotypes are drawn from the same distribution in all populations".

3.3.6 Haplotype inference

Haplotypes were inferred from genotypic data using algorithms provided by PHASE v2.1 (Stephens *et al.* 2001; Stephens and Scheet 2005) within DnaSP v5

(Librado and Rozas 2009) using the default settings of 100 main iterations, a thinning interval of 1 and 100 burn-in iterations. To check reliability of the haplotype reconstruction the algorithm was run five times for each data set and checked for consistency across the runs; no modifications of the default settings were required to improve reliability. PHASE imputes haplotypes even when genotypic data for an individual is missing; however, if more than one SNP per gene was missing for a particular individual the resulting inferred haplotype was deemed unreliable and excluded from further analyses.

3.3.7 Natural selection

For the haplotype frequency data, evidence for selection was examined by performing the Ewens-Watterson homozygosity test of neutrality (Ewens 1972; Watterson 1978) on the inferred haplotypes, and implemented in Arlequin version 3.5 (Excoffier and Lischer 2010). The test was devised by Watterson (1978) and uses the test statistic F as the sum of squared allele (haplotype) frequencies, equivalent to the sample homozygosity expected under Hardy-Weinberg equilibrium, . Significance is tested by comparing the sample homozygosity against distributions of F values expected for populations under neutrality for a given number of alleles (K) and population size ($2n$), as predicted by Ewens' sampling theory (Ewens 1972). An increased observed F value compared to that expected under neutrality is indicative of a shortage of heterozygotes; this may be due to positive selection selecting for an allele that has therefore become common or negative selection acting to keep new, deleterious alleles at low frequencies. A decreased F indicates alleles being maintained at more equal frequencies than expected, a signature of balancing selection. In Arlequin, p-values are ascertained by simulating random neutral samples having the same number of haplotypes as the real data; 50,000 random permutations were performed in this case. Ewens-Watterson tests were one-sided, based on the direction of the test statistics of the neutrality analyses performed in Chapter 2. Therefore, *Il1b*, *Il2*, *Slc11a1*, and *Tnf* were tested under the hypothesis of decreased homozygosity and balancing selection, with significance assumed if $p < 0.05$. In contrast, the genes

Il12b, *Tgfb1*, *Tlr2* and *Tlr4* were tested under the assumption of directional selection, where $p > 0.95$ would be indicative of a significant shortage of heterozygotes. To establish whether any observed departures from neutrality were down to genome-wide effects, such as those produced by demographic processes, a Mann-Whitney *U*-test was used to test the hypothesis that observed and expected homozygosity values were drawn from the same distribution. A genome-wide departure from neutrality resulting from population-level demographic changes would lead to an overall difference between observed homozygosity and that predicted under neutral expectations, whereas if selection is acting on specific genes there should be no overall difference between observed and expected values of *F* but instead only local, gene-specific departures from neutrality.

In order to detect specific SNPs that may be under selection, an outlier test was performed, based on identifying SNP loci that present F_{ST} values that are significantly different from those expected under neutrality and a given demographic model (Excoffier *et al.* 2009). Selection may affect genetic diversity between populations such that loci under local directional selection should show relatively large differences between populations, whereas loci under balancing selection across should show allele frequencies that are too even across populations (Cavalli-Sforza 1966; Lewontin and Krakauer 1973). The main difficulty of such tests is therefore to obtain the expected distribution of F_{ST} (Beaumont 2005). The test used in this study was that of Excoffier *et al.* (2009) and implemented in Arlequin v.3.5. This test generates a null distribution via coalescent simulations using a hierarchical island model, which produces fewer false-positive loci than models that do not take into account a hierarchical structure (when one exists), and is robust to uncertainties about population structure (Excoffier *et al.* 2009). The authors recommend the use of a hierarchical island model with at least twice as many groups as that suggested by the genetic structure and a large number of demes per group ($d \geq 100$); therefore the model was run with 10 groups of 100 demes, for 50,000 simulations.

3.4 RESULTS

3.4.1 Genotyping

Eighteen single nucleotide polymorphisms within eight immune genes were genotyped in a total of 814 Kielder field voles. As there appeared to be no significant difference in allele frequencies between voles sampled as part of the cross sectional or longitudinal studies at the two within-Kielder sites (those that had been genotyped from livers and tail bleeds, respectively) voles from these two groups were treated as independent observations and pooled together to increase sample size for population genetic analyses. Unless explicitly stated, all analyses were performed on the combined BLB and SQC data, as there was no evidence to suggest these two sites were separate populations (see below).

Of the 418 replicated genotypes used as control samples, 94.5% typed successfully in duplicate. Twenty-three (5.5%) samples genotyped for one sample only. No mismatches occurred between the remaining 395 duplicated genotypes, which demonstrates the reliability of the genotyping method. Furthermore, no genotypes were called from any of the negative controls; this was true of both the samples typed directly from genomic DNA and of those typed after whole genome amplification. All animals that were genotyped at the sham *Il10* SNP were correctly called as *A*:*A* homozygotes, suggesting that the assay utilised by KBioscience is robust to spurious allele calls.

3.4.2 SNP frequencies

SNP frequency data for both BLB and SQC sites combined can be seen in Table 3.2 (frequency data for the two separate subpopulations can be found in Appendix Table 1.1). Frequencies of SNP minor alleles ranged from 0.001, where a single individual was heterozygous at the Tlr2 1706 G/A SNP, to 0.37. Only one of the 18 SNP loci deviated from Hardy-Weinberg expectations: heterozygosity was lower than expected at the Tlr2 1944 T/C SNP; when examined further this was seen to be due to decreased heterozygosity at the BLB site ($p = 0.001$) but not at SQC ($p = 0.421$). The Tlr2 1706 G/A SNP was monomorphic for the *G* allele at the SQC site, and only one heterozygous individual at the BLB site harboured the *A* allele.

However, although still rare this allele is segregating at higher frequencies in other areas – the analyses of a wider region showed that the frequency of the *A* allele is 0.01 in Kielder as a whole, 0.02 in nearby Redesdale Forest and as high as 0.04 in Kershope (Appendix Table 1.2). The derived allele at the Tlr4 1663 A/G SNP is the most common and is segregating at frequencies between 0.92 and 0.99 (Appendix Tables 1 and 2). Values of the inbreeding coefficient F_{IS} were variable across the SNP loci but low overall and scattered around the mean (Figure 3.2). No F_{IS} values differed significantly from zero, bar the aforementioned Tlr2 1944 T/C SNP which exhibited an increased level of homozygosity.

Table 3.2 Population data for the 18 SNPs genotyped in the BLB and SQC populations. Results shown are overall results combined between the two sites.

Gene	SNP	*N*	MAF	H_O	H_E	HWE *p*	F_{IS}
Il1b	Il1b 243 G/A	571	0.16	0.26	0.26	0.85	0.000
	Il1b 253 A/G	568	0.32	0.43	0.44	0.75	0.025
	Il1b 324 C/T	568	0.21	0.33	0.33	0.28	0.001
Il2	Il2 381 A/T	562	0.14	0.25	0.25	0.11	-0.039
	Il2 408 C/G	567	0.29	0.41	0.41	0.19	0.005
Il12b	Il12b 278 G/C	570	0.05	0.10	0.10	0.92	0.012
	Il12b 704 C/T	569	0.02	0.03	0.03	0.25	0.082
Slc11a1	Slc11a1 537 C/G	568	0.26	0.39	0.39	0.85	-0.022
	Slc11a1 714 G/A	571	0.28	0.41	0.40	0.80	-0.022
Tgfb1	Tgfb1 442 C/A	565	0.25	0.37	0.38	0.07	0.018
Tlr2	Tlr2 1383 G/A	576	0.06	0.10	0.11	0.51	0.041
	Tlr2 1648 G/A	564	0.10	0.18	0.18	0.97	0.008
	Tlr2 1706 G/A	571	0.001	0.00	0.00	na	0.000
	Tlr2 1944 T/C	565	0.30	0.39	0.42	**0.004**	**0.070**
Tlr4	Tlr4 1663 A/G	561	0.06	0.10	0.10	0.16	0.079
	Tlr4 1848 G/T	559	0.02	0.04	0.04	1.00	-0.020
	Tlr4 2037 C/A	563	0.05	0.10	0.10	0.91	-0.018
Tnf	Tnf 210 T/C	570	0.37	0.46	0.47	0.09	0.018
		Average	0.16	0.23	0.23	0.51	0.012

N, number of individuals genotyped; MAF, minor allele frequency; H_O, observed heterozygosity; H_E, expected heterozygosity; HWE *p*, deviation from Hardy-Weinberg proportions *p*-value

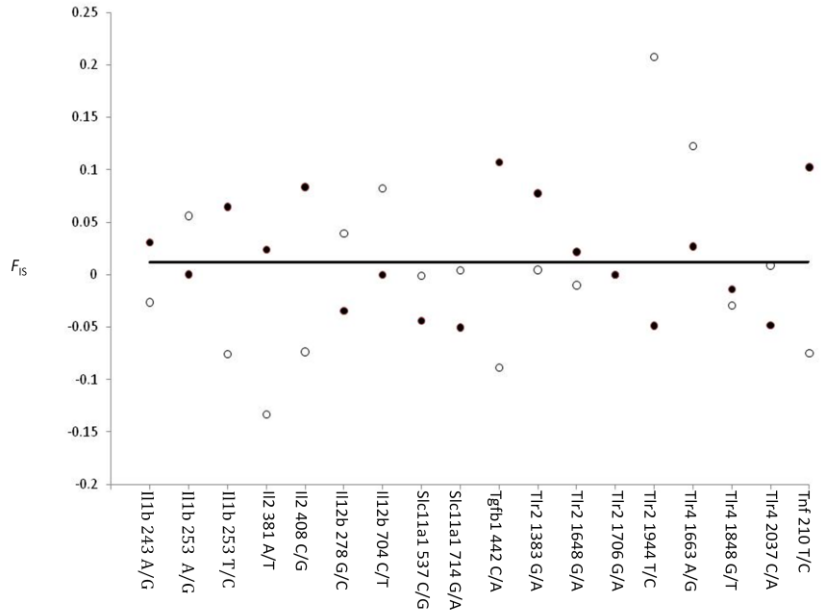

Figure 3.2 F_{IS} values plotted by SNP marker. Open circles are values for SQC site, closed circles are BLB site. Solid line represents the average F_{IS} value over all loci over both sites.

3.4.3 Linkage disequilibrium

Estimates of the extent of linkage disequilibrium were calculated from the combined SQC and BLB population genetic data. Significant LD was observed between most markers located within the same gene (Table 3.3). No significant LD was seen between the two *Il12b* SNPs, or between several pairwise comparisons involving the Tlr2 1383 G/A and Tlr2 1706 G/A SNPs with other *Tlr2* markers; however it should be noted that minor allele frequencies are low at these loci and may not give reliable estimates of LD. Significant LD was seen between other alleles within the *Tlr2* gene (Table 3.3).

Table 3.3 Pairwise linkage disequilibria between pairs of SNPs located within the same gene. Blanks, no LD coefficient returned by LinkDos when LD is non-significant.

Gene	SNP locus 1	SNP locus 2	Δ_{ij}	R_{ij}	p
Il1b	Il1b 243 G/A	Il1b 253 A/G	0.043	0.251	**0.0001**
	Il1b 243 G/A	Il1b 324 C/T	0.038	0.255	**0.0001**
	Il1b 253 A/G	Il1b 324 C/T	0.055	0.284	**0.0001**
Il2	Il2 381 A/T	Il2 408 C/G	0.041	0.262	**0.0001**
Il12b	Il12b 278 G/C	Il12b 704 C/T	-	0.019	0.650
Slc11a1	Slc11a1 537 C/G	Slc11a1 714 G/A	0.182	0.943	**0.0001**
Tlr2	Tlr2 1383 G/A	Tlr2 1648 G/A	-	0.058	0.157
	Tlr2 1383 G/A	Tlr2 1706 G/A	-	0.014	0.733
	Tlr2 1383 G/A	Tlr2 1944 T/C	0.020	0.181	**0.0001**
	Tlr2 1648 G/A	Tlr2 1706 G/A	-	0.020	0.635
	Tlr2 1648 G/A	Tlr2 1944 T/C	0.035	0.248	**0.0001**
	Tlr2 1706 G/A	Tlr2 1944 T/C	-	0.025	0.546
Tlr4	Tlr4 1663 A/G	Tlr4 1848 G/T	0.020	0.585	**0.0001**
	Tlr4 1663 A/G	Tlr4 2037 C/A	0.048	0.928	**0.0001**
	Tlr4 1848 G/T	Tlr4 2037 C/A	0.019	0.620	**0.0001**

Two pairwise comparisons between SNPs in different genes showed significant LD after correcting for multiple testing, those between Il12b 704 C/T and Tlr4 1848 G/T ($\Delta_{ij} = 0.005$, $R_{ij} = 0.263$, $p = 0.010$), and between Il2 C/G and Tlr2 1944 T/C ($\Delta_{ij} = 0.031$, $R_{ij} = 0.146$, $p = 0.045$). Ohta's analysis of the variance of LD showed that the apparent LD between these two pairs of non-linked SNPs was likely due to genetic drift between subpopulations as in both cases $D_{IS}^2 - D_{ST}^2$ was negative; -0.0007 for the comparison between Il12b 704 C/T and Tlr4 1848, and -0.0016 between the Il2 C/G and Tlr2 1944 T/C SNPs.

3.4.4 Population differentiation

Genetic differentiation as measured by overall F_{ST} between subpopulations was very low and did not suggest population subdivision. However, F_{ST} values did differ markedly between individual SNPs and there was significant genotypic differentiation at these loci (Table 3.4). The *p*-values reported for overall genotypic differentiation between sites implied that there are significant genetic differences

Table 3.4 Population subdivision between sites. For pairwise F_{ST} values, negative values approximate to zero. *P*-values reported are for tests of genotypic differentiation between populations with statistically significant results highlighted in bold. n_1, n_2 = sample size from the listed sites.

		Pairwise population comparison											
		BLB SQC			**Kielder Kershope**			**Kielder Redesdale**			**Kershope Redesdale**		
Gene	**SNP**	n_1, n_2	F_{ST}	***P-value***	n_1, n_2	F_{ST}	***P-value***	n_1, n_2	F_{ST}	***P-value***	n_1, n_2	F_{ST}	***P-value***
Il1b	Il1b 243 G/A	268, 303	0.010	**0.010**	49, 48	0.010	0.219	49, 57	-0.011	0.756	48, 57	0.002	0.305
	Il1b 253 A/G	268, 300	0.006	**0.039**	50, 48	-0.010	0.778	50, 57	-0.009	0.784	48, 57	-0.010	1.000
	Il1b 324 C/T	268, 300	0.000	0.311	50, 47	0.000	0.360	50, 56	0.006	0.206	47,56	0.038	**0.040**
Il2	Il2 381 A/T	265, 297	0.007	**0.029**	50, 47	-0.001	0.354	50, 57	-0.006	0.587	47, 57	-0.008	0.703
	Il2 408 C/G	267, 300	0.007	**0.028**	50, 48	0.001	0.311	50, 57	0.009	0.203	48, 57	-0.009	0.857
Il12b	Il12b 278 G/C	270, 300	0.011	**0.011**	50, 48	0.030	0.051	50, 57	0.010	0.165	48, 57	-0.005	0.626
	Il12b 704 C/T	270, 299	0.033	**<0.001**	50, 48	-	-	50, 56	-	-	48, 56	-	-
Slc11a1	Slc11a1 537 C/G	268, 300	0.012	**0.005**	50, 48	-0.010	0.762	50, 55	-0.009	0.877	48, 55	-0.008	0.667
	Slc11a1 714 G/A	269, 302	0.011	**0.006**	49, 48	-0.011	1.000	49, 56	-0.010	1.000	48, 56	-0.012	1.000
Tgfb1	Tgfb1 442 C/A	266, 299	0.000	0.385	49, 48	-0.003	0.432	49, 57	-0.006	0.533	48, 57	-0.010	0.881
Tlr2	Tlr2 1383 G/A	267, 309	-0.001	0.708	50, 48	0.018	0.104	50, 57	-0.006	0.732	48, 57	0.038	**0.026**
	Tlr2 1648 G/A	265, 299	-0.001	0.425	49, 48	-0.009	0.740	49, 56	-0.006	0.703	48, 56	-0.001	0.466
	Tlr2 1706 G/A	272, 299	0.000	0.476	50, 48	0.010	0.199	50, 57	-0.007	1.000	48, 57	0.001	0.409
	Tlr2 1944 T/C	269, 296	0.000	0.418	49, 46	0.004	0.288	49, 57	0.010	0.200	46, 57	-0.010	0.883
Tlr4	Tlr4 1663 A/G	267, 294	0.001	0.320	50, 48	0.051	**0.013**	50, 56	0.037	**0.034**	48, 56	-0.008	0.789
	Tlr4 1848 G/T	262, 297	0.004	0.102	50, 48	0.044	**0.024**	50, 55	0.008	0.497	48, 55	0.008	0.246
	Tlr4 2037 C/A	267, 296	-0.001	0.431	50, 48	0.051	**0.015**	50, 57	0.035	**0.034**	48, 57	-0.008	0.788
Tnf	Tnf 210 T/C	269, 301	-0.002	1.000	50, 48	-0.008	0.759	50, 57	-0.007	0.675	48, 57	-0.011	0.895
		Total	0.005	**0.001**	**Total**	0.001	**0.016**	-	-0.001	0.479	-	-0.005	0.810

between the BLB and SQC populations, and between the populations of Kielder and Kershope Forests but not between Redesdale and the two other forest sites (Table 3.4). It should be stressed however that this test assumes statistical independence across all loci, which is not the case for the linked SNPs in this study, and therefore these results must be taken as approximate (see Discussion).

3.4.5 Haplotype diversity and tests for selection

A summary of haplotype number and diversity estimates for each gene and for both sites can be seen in Table 3.5; a full outline of all the individual haplotypes can be found in Appendix Table 1.3. It should be noted that not all SNPs discovered in Chapter 2 were able to be genotyped in all genes, and that estimates of haplotype diversity for the genes *Il12b*, *Tlr2* and *Tnf* are therefore likely to be underestimates. The highest levels of haplotypic diversity were seen for *Il1b*, where H_d estimates in this larger sample were similar to those from the sequence data (Chapter 2), while *Il2*, *Tlr2* and *Tnf* also returned reasonably high estimates. Low levels of haplotypic diversity were seen for the genes *Il12b* and *Tlr4*.

Two genes, *Il1b* and *Il2*, returned significant results for the Ewens-Watterson test, both for a deficit in homozygotes (Table 3.5). This increased heterozygosity is a result of a more even allele (haplotype) frequency than that predicted from a model of neutral evolution (Figure 3.3 and Appendix Table 1.3) and can be a consequence of balancing selection. *Tnf* also demonstrated an excess of heterozygotes and was marginally non-significant ($p = 0.06$) in the test for balancing selection. The Mann-Whitney *U*-test could not reject the hypothesis of sets of observed and expected homozygosity values being drawn from the same distribution ($n_1 = 8$, $n_2 = 8$, $U = 52.0$, $p = 0.104$).

Table 3.5 Haplotype diversity and summary of Ewens-Watterson tests of neutrality for the BLB and SQC sites. Tests were one-sided; significance was assumed where $0.05 > p > 0.95$. Significant departures from neutral expectations are in bold.

Gene	Site	2*n*	N_h [a]	H_d [b]	Obs. *F* [c]	Exp. *F* [d]	*p*-value
Il1b	BLB	466	5	0.729	0.273	0.576	**0.014**
	SQC	660	5	0.725	0.276	0.591	**0.012**
	All	1126	5	0.729	0.272	0.615	**0.008**
Il2	BLB	468	3	0.571	0.430	0.739	**0.048**
	SQC	644	3	0.576	0.425	0.748	**0.041**
	All	1112	3	0.574	0.426	0.764	**0.036**
Il12b	BLB	475	3	0.179	0.821	0.739	0.573
	SQC	653	2	0.051	0.949	0.858	0.519
	All	1128	3	0.107	0.893	0.765	0.624
Slc11a1	BLB	480	4	0.474	0.527	0.649	0.344
	SQC	658	3	0.397	0.604	0.750	0.295
	All	1138	4	0.433	0.568	0.682	0.359
Tgfb1	BLB	454	2	0.454	0.548	0.852	0.094
	SQC	650	2	0.359	0.642	0.859	0.167
	All	1104	2	0.404	0.596	0.869	0.123
Tlr2	BLB	468	5	0.602	0.400	0.575	0.195
	SQC	660	4	0.617	0.384	0.661	0.071
	All	1128	5	0.611	0.390	0.614	0.133
Tlr4	BLB	458	5	0.153	0.848	0.575	0.883
	SQC	648	3	0.092	0.908	0.748	0.693
	All	1106	5	0.118	0.882	0.612	0.870
Tnf	BLB	470	2	0.481	0.520	0.852	0.061
	SQC	650	2	0.471	0.530	0.859	0.068
	All	1120	2	0.475	0.525	0.869	0.060
		Average (SD)	3.6 (1.3)	0.43 (0.22)	0.57 (0.22)	0.72 (0.11)	0.27

[a] Number of haplotypes

[b] Haplotype diversity

[c] Observed values of Watterson's test statistic *F*

[d] Expected values of *F* under neutrality given the population size and number of haplotypes (Ewens 1972)

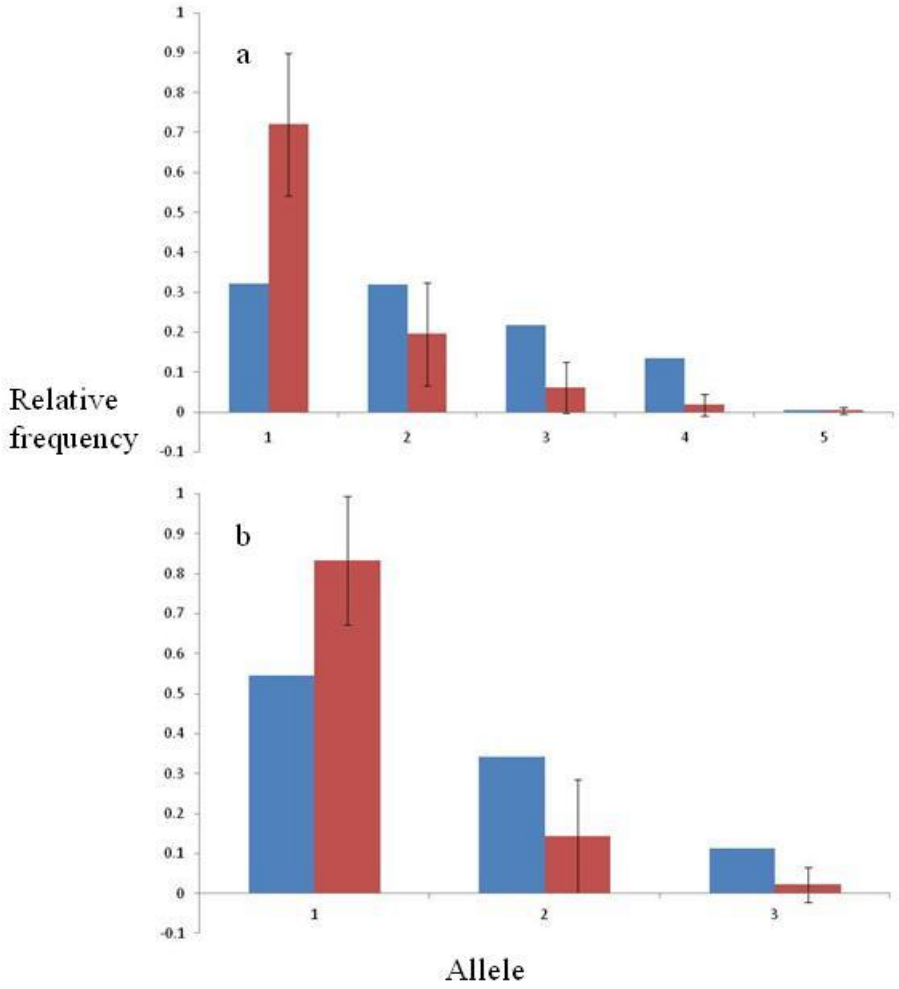

Figure 3.3 Observed (blue) and expected (red) allele frequencies for the genes (a) *Il1b* and (b) *Il2*. Error bars represent 1 s.d. Allele frequencies are more even than expected under neutrality.

The Ewens-Watterson test was also performed on the data from the wider Kielder, Redesdale and Kershope Forest samples (Appendix Table 1.4). *Il1b* was again significant for evidence of balancing selection, displaying decreased homozygosity and a more even-allele frequency than expected under neutrality, within Kielder (p = 0.044) and Redesdale forests (p = 0.003), and over all three sites combined (p = 0.024). The *Il1b* genotyping data from Kershope Forest was marginally non-significant (p = 0.066). The genes *Il2* and *Tnf* were marginally non-significant for balancing selection in this smaller and geographically more wide-ranging dataset (p = 0.090 and p = 0.079, respectively). For the data from Kielder Forest, *Tlr4* had a much higher level of homozygosity than expected under neutral evolution

(observed $F = 0.98$, expected $F = 0.81$, $p = 1.000$; Appendix Table 1.4), which may be indicative of directional selection. This pattern was not repeated within either Kershope or Redesdale forests. An additional Mann-Whitney *U*-test was performed on this dataset which again could not reject the hypothesis of observed and expected values of *F* being drawn from the same distribution $n_1 = 8$, $n_2 = 8$, $U = 52.0$, $p = 0.227$).

The F_{ST}-based outlier analysis predicted four individual SNP loci to possibly be under selection: Il1b 324 C/T, Il12b 704 C/T, Tlr2 1706 G/A and Tlr4 1663 A/G. The Il12b 704 C/T and Tlr2 1706 G/A SNPs were monomorphic at the SQC site and this may be what the test has detected and interpreted as a large difference between the two sites. The Il1b 324 C/T SNP has a much lower F_{ST} value than the neighbouring, linked SNPs within the *Il1b* locus. The Tlr4 1663 A/G SNP, at which the derived allele is segregating at frequencies close to fixation, demonstrated the highest F_{ST} values between Kielder and the two other forests, but low F_{ST} between the two within-Kielder sites.

3.5 DISCUSSION

In order to gain a more complete understanding of how natural selection shapes genetic variation it is important to examine patterns of genetic variation within non-human, natural populations. Very few studies have attempted to characterise the role of natural selection in maintaining population genetic diversity in non-MHC immune genes of wildlife (Acevedo-Whitehouse and Cunningham 2006; Tonteri *et al.* 2010). This chapter builds on the estimates of DNA sequence diversity in the previous chapter by examining genetic diversity at the population level. I have found extensive variation at immune gene loci within natural field vole populations and provided further evidence for natural selection shaping the patterns of diversity seen within some of these genes.

3.5.1 SNP frequencies and population genetics

No widespread departures from Hardy-Weinberg equilibrium were found; only one SNP in one subpopulation deviated from Hardy-Weinberg proportions. Inbreeding was likely to be low, due to the low overall F_{IS} value and the wide variation in the inbreeding coefficient among individual SNPs; inbreeding would lead to a genome-wide effect and increased homozygosity across the whole genome (Black *et al.* 2001), which was not the case in this study.

Pairwise F_{ST} values between the two within-Kielder subpopulations were very low, suggesting little genetic difference between the two populations. However, tests of genotypic differentiation implied that there was a significant genetic difference at various individual SNP loci, and overall between these populations. It should be noted however that the estimates of population structure in this chapter will be approximate. Neutral genetic markers such as microsatellites are better suited to ascribe the relative contributions of genetic drift and migration to observed levels of genetic diversity, as neutral diversity is governed by these forces alone (Piertney and Webster 2010); however, variation at neutral loci cannot provide direct information on selective processes (Sommer 2005). My aim was to understand how selection has shaped the level of diversity observed in field voles and the consequences of this adaptive variation. Although overall genetic differentiation between subpopulations was low, F_{ST} values at individual loci differed markedly, which may be indicative of natural selection acting on specific genes or individual SNPs. I do however acknowledge that the use of a panel of neutral markers would have been an advantage in this study, both in terms of enabling a more accurate estimation of genetic differences between populations and insights into the mechanisms behind such variation, and to allow comparisons with the putatively non-neutral SNPs analysed in this study.

Linkage disequilibrium was hypothesised to be strong within genes, due to linkage between segregating sites, but not present between SNP markers located in different genes, which allows for each gene to be treated as independent. Overall, this pattern was indeed seen. Linkage disequilibrium within a gene is important for detecting

associations with phenotypic variation, particularly if the causal mutation is up- or downstream from the SNPs that have been genotyped; this will become particularly relevant in later chapters which investigate associations between genetic markers and phenotypic traits. Strong within-gene LD such as that found here is also important for robust haplotype inference.

3.5.2 Natural selection

The results of this chapter support the findings of the previous, in that those genes that exhibited signs of natural selection at the DNA sequence level also did so at the population level. As there was no suggestion of genome-wide departures from neutrality, it is likely that the patterns of diversity at those genes exhibiting departures from neutrality are due to natural selection, as selection acts locally on a few genes whereas demography or inbreeding would affect all loci similarly across the genome.

Evidence for balancing selection acting to maintain diversity at *Il1b* and *Il2* was once again observed in this chapter. In Chapter 2 these genes displayed an excess of intermediate frequency polymorphisms; here, over several subpopulations an excess of heterozygotes and a more even frequency of haplotypes than one would expect under neutrality provided further evidence for balancing selection at these loci. An *Il1b* homologue has previously been identified as a possible candidate for selection in Atlantic salmon (Tonteri *et al.* 2010), while interleukins in general have previously been identified as targets of balancing selection in humans, with parasitic worms thought to act as the major selective force (Fumagalli *et al.* 2009).

In the previous chapter *Tnf* exhibited evidence for balancing selection across multiple tests of neutrality; in this chapter *Tnf* was marginally non-significant for the Ewens-Watterson test at the population level. However, it should be noted that of the two SNPs discovered within the *Tnf* locus and analysed in Chapter 2, only one was able to be genotyped and incorporated into the population-level analyses of this chapter. It is likely therefore that the haplotype frequency and diversity estimates for *Tnf* are less accurate than they otherwise would have been, and that

the Ewens-Watterson test therefore had less power to identify a signature of balancing selection. Overall, the fact that multiple tests rejected neutrality for the genes *Il1b*, *Il2* and *Tnf*, coupled with unusually high diversity estimates and a lack of support for genome-wide effects, provides strong evidence for balancing selection acting at these genes. In humans, balancing selection is a relatively common mechanism for maintaining adaptive variation in immune genes (Ferrer-Admetlla *et al.* 2008). Patterns of genetic diversity consistent with balancing selection have also been observed previously in the wild across several families of immunity genes, including the MHC (Piertney and Oliver 2006), antimicrobial peptides (Tennessen and Blouin 2008) and *TAP* genes (Transporter associated with Antigen Processing) (Jensen *et al.* 2008), while cytokine genes have exhibited signs of balancing selection in humans and chicken (Fumagalli *et al.* 2009; Downing *et al.* 2010).

In the previous chapter *Tlr4* was close to significance for signs of directional selection over several tests of neutrality. Although these tests were marginally non-significant ($0.05 < p < 0.06$) the consistency across several tests may indicate that natural selection has acted on *Tlr4*. The analyses of this chapter revealed further signs that directional selection may have played a role in shaping the variation within this gene, and at the nonsynonymous Tlr4 1663 A/G SNP in particular; at this site the derived *A* allele is segregating at very high frequencies, which may be indicative of selection driving the increase in frequency of this relatively recent allele in field voles. In addition, this individual SNP was identified as a candidate for selection by the F_{ST}-based outlier test. The *A* allele of the Tlr4 1663 A/G SNP is segregating at high frequencies in all populations, but particularly so in Kielder Forest (99%). The Kielder population was also consistent for signs of positive selection when analysing haplotype frequencies using the Ewens-Watterson test. The frequency of the most common AGC haplotype is 0.99, which is much higher than that expected under neutral evolution. Haplotypes containing the ancestral *G* allele at the Tlr4 1663 A/G SNP are at very low frequencies. A possible explanation for these patterns may be that the *A* allele offers protection against a pathogen or pathogens which are prevalent at higher frequencies within Kielder

Forest, resulting in greater selective pressures for this allele within Kielder compared to surrounding areas.

Deviations from neutral evolution can also result from demographic processes; for example, an excess of intermediate frequency mutations, like those observed here for the genes *Il1b*, *Il2* and *Tnf*, may be indicative of balancing selection but can also indicate population subdivision or contraction, as rare alleles are lost during population contractions (Nachman 2006). The task of disentangling the effects of demography and identifying specific genes under selection can be challenging (Beaumont and Balding 2004). Kielder field vole populations are known to exhibit multi-annual density cycles of crashes followed by periods or recovery (Lambin *et al.* 2000) and these demographic processes could affect patterns of genetic diversity. However, the population declines associated with these cycles are relatively short (1 -2 years; Lambin *et al.* 2000) in comparison with evolutionary timescales. Furthermore, the population of field voles within Kielder Forest as a whole (i.e. the metapopulation) is likely to have a relatively constant effective population size since, first, cycles among subpopulations in Kielder Forest are not synchronous (Lambin *et al.* 1998) and, second, migration is likely to be frequent between subpopulations. Taken together, this would argue against a demographic (population contraction) explanation of the patterns of genetic variation observed.

3.5.3 Parasites as a selective force

Parasite-driven selection represents a major driving force underpinning the evolution of immune genes (Sommer 2005; Barreiro and Quintana-Murci 2010) and indeed has been implicated in the evolution of cytokine genes in humans (Fumagalli *et al.* 2009). As Kielder field voles are known to harbour a wide variety of pathogens, endemic parasites represent a possible selective force driving the patterns of immunogenetic variation observed in these animals. Various mechanisms for pathogen-driven balancing selection acting at immune gene loci have been proposed, including overdominant selection, or heterozygote advantage, where the relative fitness of heterozygotes is consistently higher than homozygotes (Doherty and Zinkernagel 1975), and negative frequency-dependent selection, or

rare-allele advantage, where the fitness of an allele is inversely proportional to its frequency (Bodmer 1972; Takahata and Nei 1990). A third model (Hedrick 2002) suggests that fluctuating selection, occurring where the fitnesses of alleles vary in time or space due to the variation in the prevalence of pathogens or other selective agents (Tennessen and Blouin 2008; Bell 2010), is sufficient to maintain polymorphism at immune genes. Fluctuating selection effectively amounts to directional selection for different alleles at different time points or in different environments. As spatial and temporal variation in parasite prevalence within Kielder has been observed both in this study (Chapter 4) and previously (Cavanagh *et al.* 2004; Smith *et al.* 2005; Telfer *et al.* 2007a; Telfer *et al.* 2007b; Burthe *et al.* 2008a), fluctuating selection and local adaptation to parasites is a plausible explanatory mechanism for the balanced diversity observed at the *Il1b* and *Il2* loci, at which allele frequencies differed between neighbouring subpopulations.

Natural selection for increased resistance to a pathogen may lead to an increase in frequency for alleles that are otherwise detrimental (Dean *et al.* 2002); for example, a particular host genotype may protect an individual following exposure to an infection but in the absence of infection may carry a fitness cost, such as reduced fecundity or viability. Such antagonistic effects on fitness may be particularly likely to occur for immune gene polymorphisms as some immune responses to infection, such as inflammation, can be harmful to host tissues. For example, it is known in humans that the magnitude of cytokine-mediated inflammatory responses has a genetic basis, and while an effective immune response is crucial for the clearance of infection, too strong a response can lead to significant immunopathology to the host (Graham *et al.* 2005). Alternatively, a genotype that confers resistance to one infectious disease may also increase susceptibility to another, or to a non-communicable disease (Stearns 1999). Therefore it is conceivable that where parasite prevalence differs in time and/or space, as it does within Kielder Forest, different alleles may be selected for at different times or locations.

In order to elucidate the precise mechanism by which genetic diversity is maintained in field vole immune genes it is important to understand how the observed genetic diversity is associated with variation in phenotypic traits, as functional variation is a prerequisite for the action of natural selection. In order to verify that parasites are indeed the driving force of selection at these loci, it is essential therefore to link the genotypes of individuals with susceptibility to naturally-occurring parasites; this will be the aim of the subsequent chapters.

3.6 CONCLUSION

Understanding how genetic variation is shaped by selection in natural populations is of fundamental importance in evolutionary biology. This study is the first to examine the genetic diversity of a wide range of cytokine and other immune genes in a natural population. Strong evidence for balancing selection acting on several of these genes (*Il1b*, *Il2* and *Tnf*), alongside support for positive selection at *Tlr4*, has been observed at both the population and DNA sequence level.

Subsequent chapters will examine genetic associations with parasite resistance and gene expression in an attempt to elucidate the mechanisms of how and why selection is acting to maintain diversity at cytokine gene loci in the field vole.

Chapter 4

Immune gene polymorphism and parasite resistance

4.1 INTRODUCTION

A primary aim of candidate gene studies is to link genetic polymorphism to phenotypic variation and, ultimately, fitness. Parasites (including viruses, bacteria, protozoa, helminths and arthropods) can have a large impact on host fitness and are a major cause of morbidity and mortality in natural populations, with the potential to drive threatened populations to extinction (Smith *et al.* 2009). A better understanding of the genetic basis of susceptibility to parasites within natural populations has the potential to enable the identification of at-risk individuals or populations and to uncover novel mechanisms associated with infectious disease susceptibility and pathogenesis. Therefore, such work has important implications for conservation biology, disease control and epidemiology (Cleaveland *et al.* 2002; Acevedo-Whitehouse and Cunningham 2006), which may currently be particularly relevant as climate change drives the increased emergence of novel pathogens in natural populations (Harvell *et al.* 2002).

The vast majority of studies into the association between immunogenetic variation and parasitic disease in wildlife have focussed on genes of the major histocompatibility complex (reviewed in Piertney and Oliver 2006). Although MHC genes are undoubtedly important, their contribution to the immune response is relatively small, as the cumulative contribution of non-MHC genes exceeds that of the MHC (Jepson *et al.* 1997). By way of an example, Behnke *et al.* (2003) found that non-MHC genes are far more important for resistance to the nematode *Heligmosomoides polygyrus* in laboratory mice than MHC genes. There are, then, a number of non-MHC genes that contribute significantly to the immune response and which are therefore likely to be associated with infectious disease susceptibility, including cytokines and TLRs. Indeed, each of the genes examined in this chapter are known to have important roles in the immune response and several have been

implicated in variation in susceptibility to - or progression of - infectious diseases in humans or model organisms (Romani *et al.* 1997; Reed 1999; Van Deventer 2000; Blackwell *et al.* 2001; Bayley *et al.* 2004; Schröder and Schumann 2005; Gomez *et al.* 2006; Smith and Humphries 2008). However, to my knowledge, only one study thus far has attempted to link polymorphism within cytokines or other non-MHC immune genes to infectious disease resistance within a natural population: Coltman *et al*. (2001) found an association between polymorphism at an *Ifng*-linked microsatellite and resistance to gastrointestinal nematodes in free-living Soay sheep.

Previous chapters have identified widespread variation within immune genes of field voles. The functional relevance of this diversity has been inferred from accumulated evidence demonstrating that the patterns of diversity observed within several of these genes have been shaped by natural selection, particularly balancing selection. Many endemic pathogens are present in Kielder, including but not limited to those described in this chapter, and I have hypothesised that the spatiotemporal variation in these naturally-occurring parasites may be the selective force driving the observed patterns of genetic diversity. If parasites are indeed providing the adaptive pressure that shapes and maintains diversity at these loci, then polymorphism within these selected genes should be associated with variation in infectious disease resistance.

4.2 OUTLINE

The aim of this chapter is to examine the relationship between immune gene polymorphism and phenotypic variation in pathogen resistance, in order to examine the hypothesis that parasites are the selective force driving the maintenance of diversity observed at several field vole genes. Genetic associations with a range of parasites are examined, including bacteria, protozoa, helminths and ectoparasites. I propose that if contemporary parasite-driven selection is acting on these loci then these genes should be associated with variation in resistance to the naturally-occurring parasites present in Kielder Forest.

4.3 MATERIALS AND METHODS

4.3.1 Sampling site and procedure

Field voles were caught and processed as part of two ongoing studies into the immunodynamics of natural populations; one cross-sectional study involving destructive sampling of field voles (Jackson *et al.* in review) and one longitudinal capture-recapture study (Jackson *et al*., in prep.) running parallel to this. Briefly, two grassy clear-cut sites within Kielder Forest, designated SQC (55.2549, -2.6116) and BLB (55.2457, -2.6108), were used for sampling monthly between February and November 2008, and again from February to March 2009 for the cross-sectional study. Curvilinear transects of 100 Ugglan special live capture traps (Grahnab, Sweden), arranged at 5-10m intervals, were placed around the margins of each habitat in order to sample a very large area of the habitat and provide data representative of the whole clear-cut population. In addition, a rectangular 0.375 ha, 150-trap (10 × 15) live-trapping grid was placed centrally in each habitat and used in the parallel capture-recapture study running from March-October 2008. In this longitudinal study individual voles were tagged by injecting a subcutaneous Passive Induced Transponder (PIT) tag (AVID plc., Uckfield, UK) under the skin at the back of the neck and subsequently identified via their unique nine-digit code with hand-held scanners (AVID). In November and March 2009 larger samples of animals were captured and killed from both the transect and central grid habitats at both sites, including animals that had been previously marked with transponders and processed as part of the capture-recapture study. On capture, each animal was weighed, sexed and examined for ectoparasites; each vole was thoroughly examined for ticks while fleas were counted by combing each individual over water. Counts of three species of flea (*Peromyscopsylla spectabilis*, *Ctenophthalmus nobilis vulgaris*, *Megabothris walkeri*) were grouped as a single 'flea' variable in order to simplify further analyses, under the assumption that they have a similar effect on the host. Reproductive maturity was also ascertained; males were classed as sexually mature if they possessed an adult coat and showed external signs of descended testes, while females were designated as mature if they had an adult coat along with a perforate vagina and/or an open pubic symphysis and evidence of lactation.

4.3.2 Endoparasite diagnoses

Parasite diagnoses were performed by Malgorzata Zawadzka, Department of Veterinary Pathology, University of Liverpool. Each vole caught as part of the cross-sectional study was killed by an overdose of chloroform followed by exsanguination. Each animal was then dissected and examined for infection by nematodes and cestodes, recorded as presence/absence of infection. In addition, the total numbers of cestodes were counted for each infected vole. Animals caught as part of the longitudinal study were sampled non-destructively via tail bleeds (for methods see Chapter 3, section 3.3.2) and could not therefore be examined for helminths, with the exception of those animals captured and killed in November 2008 and March 2009.

Field voles from both studies were tested for infection by the blood-borne parasites *Babesia microti* and *Bartonella* spp. *Babesia microti* is an intraerythrocytic protozoan parasite infecting wild rodents, and the major causative agent of human babesiosis in the USA, a potentially fatal tick-borne zoonosis. In Europe, babesiosis is usually caused by *Babesia divergens*, an important parasite of cattle, which although rarer is more lethal (Kjemtrup and Conrad 2000). The first reported European case of human babesiosis caused by *Babeisa microti* was published on a Slovenian patient co-infected with the Lyme disease organism, *Borrelia burgdorferi* (Meer-Scherrer *et al.* 2004). Currently there is some controversy over the taxonomic status of *B. microti*; traditionally it has been grouped within the genus *Babesia*, although classical taxonomic criteria places it instead in the *Theileria* genus (Uilenberg 2006). However, molecular evidence suggests that this parasite differs both from *Babesia* and *Theileria* and that a new genus containing several species may be required (Uilenberg 2006; Nakajima *et al.* 2009). As this debate has not yet been resolved, this parasite will continue to be referred to by its traditional name of *Babesia microti* for the purposes of this study. Diagnosis of *B. microti* infection was achieved using a previously published real-time PCR method to detect *B. microti* DNA (Bown *et al.* 2008), following extraction of DNA from field vole blood samples by alkaline digestion (Bown *et al.* 2003).

Bartonella species are flea-transmitted gram-negative bacteria which infect red blood cells of a wide range of mammalian species. Diagnosis of *Bartonella* was achieved by firstly performing a total DNA extraction from blood by alkaline digestion (Bown *et al.* 2003), followed by a genus-specific PCR protocol for the detection of *Bartonella* DNA (Telfer *et al.* 2005). Although up to five species of *Bartonella* circulate concurrently in rodent communities in the UK (Birtles *et al.* 2001; Telfer *et al.* 2007a; Telfer *et al.* 2007b), all species are thought to have similar life-cycles and are predicted to interact with the host immune system in a similar way (Oliver *et al.* 2009); therefore, individual voles were recorded as either positive or negative for the presence of *Bartonella*.

4.3.3 Genotyping

Genomic DNA from field voles from the cross-sectional study was extracted from liver samples of freshly dissected animals using the DNeasy Blood and Tissue Kit (QIAGEN); animals from the longitudinal study had their gDNA extracted from blood samples using DNAzol BD Reagent (Invitrogen). Eighteen SNPs within eight immune genes were genotyped by KBioscience (for full details see Chapter 3, section 3.3.3).

4.3.4 Preliminary analyses

Several preliminary analyses were performed on the genotyping data, as errors in genotyping can markedly affect the biological conclusions of a study (Pompanon *et al.* 2005). These included testing for Hardy-Weinberg and linkage equilibria, the use of positive and negative genotyping controls and checking for unusual patterns of missing data (see Chapter 3 for full methods of preliminary analyses).

Haplotypes were inferred from genotypic data (Chapter 3). There are several reasons why studying variation in haplotypes rather than individual SNPs should be an improvement for association analyses such as the present study. First, the functional and biologically relevant product of a gene, the protein, consists of chains of amino acids whose sequences correspond to haplotypes inherited from each parent (Clark 2004). Second, variation in populations is structured into

haplotype blocks which are themselves shaped by evolutionary processes and are likely to be transmitted as a single unit (Clark 2004; Yang *et al.* 2008). Third, utilising haplotypes comprising of multiple SNPs reduces the number of tests performed in comparison to testing each individual SNP. Finally, it has been recognised that haplotype-based studies are more powerful for detecting phenotypic associations resulting from untyped causal mutations in LD with genotyped SNPs/haplotypes (Vasemägi and Primmer 2005; Yang *et al.* 2008) .

4.3.5 Statistical analysis

Statistical modelling was performed using R version 2.10 (R Development Core Team 2009). The primary aim of the analyses was to examine the relationship between genetic variation and parasite infection, whilst controlling for confounding non-genetic factors which may influence the probability or intensity of infection. The common method for each analysis, for both cross-sectional and longitudinal data, was therefore to first construct a minimal model containing only non-genetic terms. To this minimal, non-genetic model each genetic term was added in turn to assess their significance in explaining the variance of a given parasite response variable. Table 4.1 lists the various parasite response variables examined in this chapter.

Cross-sectional study

General linear models (GLMs) were used to analyse any genetic association with parasite resistance for the cross-sectional data. In all cases the parasite response variables had error structures that were not normally distributed. As such, models were fitted to either a binomial error structure with a logit-link function (for binary, presence/absence variables) or to quasi-Poisson errors with a log-link function (for count data which exhibited overdispersion). Initial, full non-genetic models were constructed for each response variable, containing the following terms as main effects and their two way interactions: site (two levels: SQC and BLB); sex (2 levels: male and female); season [five levels, designated as: spring 2008 (March to May 2008), summer 2008 (June to August 2008), autumn 2008 (September to November 2008), winter 2008 (December 2008 to February 2009) and spring 2009

Table 4.1 Parasite response variables

Variable	Type	Levels
Nematode infection	Binary	Presence Absence
Cestode infection	Binary	Presence Absence
Cestode burden	Count	-
Babesia microti infection	Binary	Presence Absence
Bartonella infection	Binary	Presence Absence
Flea infection	Binary	Presence Absence
Flea burden	Count	-
Tick infection	Binary	Presence Absence
Tick burden	Count	-

(March 2009)]; body weight (continuous) and eye-lens weight (continuous). The two weight measurements were correlated; however, both terms were retained since body weight was used as a measure of current body size and general condition, while eye-lens weight was used as a proxy for age, as weight fluctuations will not affect the size of the eye-lens. The significance of each non-genetic variable was assessed by comparing the deviance explained by the full model to that of a model which did not contain the term. The full non-genetic model was reduced by removing each non-significant term, beginning with interactions and terms that had the highest *p*-value; terms were deletion tested such that those that did not cause a significant increase in deviance when removed were left out of the model. Significance was determined by comparing the change in deviance resulting from deletion of a given term against a χ^2 distribution with degrees of freedom equal to the number of terms or levels of a factor dropped (Paterson *et al.* 1998). Following deletion testing only significant terms were left in the resulting non-genetic minimal model.

After construction of the non-genetic models, significance of genetic terms was calculated by addition of these terms to the minimal non-genetic model. Analysis of deviance tests were then used to assess whether the model including genetic terms was a significantly better fit to the data than the model without. Genetic terms were fitted to the minimal models under either a heterozygosity model, where, for each gene, values of heterozygotes were compared to homozygotes, or an additive model. The additive model compared the most common allele (haplotype) at a particular locus to all other alleles; therefore, for a specific allele at a given locus the value of the heterozygote class lies between the two homozygote classes (that is, given the comparison of an allele A_i against the most common allele, termed A_0, the genotypes A_0A_0, A_0A_i and A_iA_i have the genotype values 0, α_i and $2\alpha_i$, respectively, where α_i represents the additive effect of the A_i allele) (Paterson *et al.* 1998).

Longitudinal study

General linear mixed models (GLMMs) were used to examine genetic associations with parasite infection for the longitudinal dataset. As all parasite variables in this dataset were binary responses, a logit link and binomial errors were assumed. In all cases the GLMMs were fitted by means of the Laplace approximation to restricted maximum likelihood (REML) estimation, using the lmer procedure in R. As with cross-sectional data, analyses of longitudinal data need to account for confounding environmental and other factors which may be associated with probability or burden of infection. As such, when constructing the initial full non-genetic models I considered the following extrinsic and intrinsic factors that may have influenced susceptibility to a given parasite: site; sex; season [spring 2008 (March to May 2008), summer 2008 (June to August 2008), autumn 2008 (September to November 2008), winter 2008 (December 2008 to February 2009)]; maturity status (two levels, immature and currently/previously reproductively active); body weight (standardised by subtraction of the mean weight from each individual value); and recapture status (whether or not the animal had been caught previously). Two-way interactions were considered for all terms except recapture status. As animals caught at the same site in the same trap session experienced the same environmental

conditions, *site*trap session* was included as a random effect (Telfer *et al.* 2008). To account for the correlation amongst different observations of the same individual (pseudoreplication), *individual* was added as a random effect to the non-genetic model. However, due to problems with model convergence, *individual* was not able to be added to the model of *Babesia microti* infection.

The Akaike Information Criterion (AIC) index (Akaike 1973) was used to select the most parsimonious non-genetic minimal model. Models can be considered as similar in their ability to account for the data if they have a difference in AIC of less than two (Johnson and Omland 2004). Terms were removed from the maximal model one at a time, beginning with interactions; at each step the term that, on removal, caused the greatest reduction in AIC (or the smallest increase in AIC still less than 2 greater than the model with the lowest AIC) was left out of the model. The minimal non-genetic model was arrived at when no further terms could be removed without causing an increase in AIC of greater than two. After selection of the best non-genetic base model, genetic terms were added as before under additive and heterozygosity models. Significance of genetic terms was assumed if addition of these terms caused a reduction in AIC of greater than 2, compared to the base model. Alongside each GLMM I also constructed equivalent GLMs as a check for consistency (i.e. that if a significant association had been observed in a mixed model it should also be observed in the GLM) and to ensure that use of mixed models and addition of random effects had not led to problems with model convergence.

4.4 RESULTS

For the cross-sectional study, 307 field voles were captured and analysed for parasite infection, 152 from BLB and 155 from SQC. The longitudinal data consisted of 576 captures of 349 individuals (BLB, $n = 158$; SQC, $n = 191$), with each animal caught on average 1.7 times. The primary focus of this study was to identify genetic associations with parasite infection; the purpose of the minimal non-genetic models was to account for confounding environmental and intrinsic

factors prior to the addition of genetic terms as explanatory variables. Elucidating the precise impact of non-genetic factors on parasite resistance was not a priority of this study and will be examined more thoroughly elsewhere (Jackson *et al.*, in prep.); therefore, such factors will only be briefly discussed here. A full summary of non-genetic terms significantly associated with infection by the various parasites can be found in Appendix Tables 2.1 and 2.2 for the cross-sectional data, and Appendix Tables 2.3 and 2.4 for the longitudinal analyses. In the longitudinal dataset, where GLMMs and GLMs were constructed for each parasite response term, the results of the two methods were broadly similar, in that significant genetic associations observed using mixed models ($\Delta AIC > 2$) were also seen using GLMs ($p < 0.05$).

4.4.1 Nematodes

Nematode infection was a binary response variable recorded as either presence or absence. In total, 66 (21.5%) of the examined voles were recorded as infected. The minimal non-genetic model was defined as: nematode infection ~ site + season + sex. Male voles were more likely to harbour a nematode infection than females, while individuals caught at the BLB site had an increased probability of infection compared to those captured at SQC. There was also a seasonal effect on probability of nematode infection, with spring 2008 associated with the highest chance of being infected (Appendix Table 2.1).

Genetic terms were tested at each of the eight loci, under both additive and heterozygosity models. Addition of *Il2* as a genetic term significantly improved the fit of the model under both heterozygosity ($p = 0.044$) and additive models ($p = 0.043$). The coefficients for significant genetic terms in the nematode infection model are summarised in Table 4.2. Heterozygotes at the *Il2* locus were more likely to harbour a nematode infection than homozygotes, while voles with the rarest TC haplotype had an increased chance of nematode infection compared to the most common, AC. *Il12b* was also associated with variation in probability of

Table 4.2 Genetic terms significantly associated with parasite infection.

Gene	Response	Comparisons	Dataset [a]	Coefficient (s.e.)	*t*-value	*z*-value	*p*-value	ΔAIC [b]
Il1b	Tick infection	Heterozygote vs. homozygote	CS	0.91 (0.41)	-	2.21	0.027	-
	Bartonella infection	Heterozygote vs. homozygote	Long.	-0.61 (0.23)	-	-2.64	0.008	4.6
	Flea infection	Heterozygote vs. homozygote	Long.	-0.50 (0.24)	-	-2.10	0.036	2.1
		GGC vs. GAC	Long.	0.36 (0.19)	-	1.84	0.065	2.2
Il2	Nematode infection	Heterozygote vs. homozygote	CS	0.61 (0.31)	-	1.99	0.047	-
		TC vs. AC	CS	0.70 (0.29)	-	2.43	0.014	-
	Cestode burden	Heterozygote vs. homozygote	CS	-0.45 (0.14)	-3.25	-	0.001	-
	Flea burden	Heterozygote vs. homozygote	CS	-0.27 (0.14)	-1.98	-	0.049	-
	Tick infection	TC vs. AC	Long.	-0.66 (0.34)	-	-1.92	0.055	2.3
	Tick burden	Heterozygote vs. homozygote	CS	0.58 (0.28)	2.04	-	0.042	-
	Babesia infection	Heterozygote vs. homozygote	Long.	0.66 (0.23)	-	2.82	0.005	7.2
	Bartonella infection	Heterozygote vs. homozygote	Long.	-0.61 (0.21)	-	-2.88	0.004	5.9
Il12b	Nematode infection	Heterozygote vs. homozygote	CS	1.11 (0.41)	-	2.73	0.006	-
		CC vs. GC	CS	1.42 (0.44)	-	3.23	0.001	-
	Babesia infection	CC vs. GC	Long.	1.02 (0.31)	-	3.28	0.001	6.8
Slc11a1	Flea burden	Heterozygote vs. homozygote	CS	-0.28 (0.14)	-1.97	-	0.049	-
	Tick burden	Heterozygote vs. homozygote	CS	0.67 (0.28)	2.41	-	0.017	-
Tlr2	Cestode burden	Heterozygote vs. homozygote	CS	0.37 (0.14)	2.64	-	0.009	-
Tnf	Cestode infection	T vs. C	CS	-0.68 (0.22)	-	3.12	0.002	-

[a] Refers to cross-sectional (CS) or longitudinal (Long.) studies.

[b] Increment in AIC of the model if single term is dropped (GLMMs only)

nematode infection under both heterozygosity ($p = 0.007$) and additive models ($p = 0.006$); heterozygosity increased probability of infection, as did the CC haplotype (Table 4.2). Marginally non-significant associations between the *Slc11a1* locus and nematode infection were observed under both the additive ($p = 0.07$) and heterozygosity models ($p = 0.07$).

4.4.2 Cestodes

Data were available for cestodes to allow two analyses; one, as for nematodes, based on a binary presence/absence response variable and another based on the total number of cestodes harboured by an individual.

Cestode infection

In total, 149 (48.5%) voles were infected with cestodes in the cross-sectional dataset. The minimal non-genetic model explaining cestode infection was defined as: cestode infection ~ season + body weight + eye lens weight + sex + season × body weight. Eye lens weight was very highly negatively associated with cestode infection and males were more likely to be infected than females (Appendix Table 2.1). There was a significant interaction between season and weight; the mean weight of infected animals was higher than uninfected animals in spring, summer and autumn 2008, particularly so in summer, while in winter 2008 and spring 2009 there was no difference in the weight of infected and uninfected voles.

A genetic effect was observed for the *Tnf* locus under an additive model ($p = 0.001$) but not a heterozygosity model. Voles carrying the *Tnf* T haplotype had a decreased probability of infection by cestodes (Table 4.2).

Cestode burden

Cestode burden ranged from 0 to 21 per individual (mean = 1.4) and had an extremely skewed distribution, with 158 (51.5%) individuals having zero counts and a small number having relatively high values ($n = 7$ (2%), range = 10-21). The minimal non-genetic model was defined as follows: cestode burden ~ site + season + body weight + eye lens weight + sex + site × season + site × body weight + site ×

eye lens weight + season × weight + eye lens weight × sex (Appendix Table 2.2). The site × season interaction appears to be explained by the mean cestode intensity at BLB site (6.1) being much higher than SQC site (2.6) in summer 2008, as there is little difference at other times.

There were strong associations between cestode burden and two immune genes, *Il2* (p = 0.001) and *Tlr2* (p = 0.008), under a heterozygosity model. *Il2* heterozygotes harboured significantly fewer cestodes than homozygotes, while heterozygosity at the *Tlr2* locus was associated with an increased cestode intensity (Table 4.2).

4.4.3 Fleas

Flea infection

Just over half of the captured animals in the cross-sectional study were infected by fleas (n = 156; 50.8%). The base model examining an individual's probability of flea infection was: flea infection ~ site + season + weight + site × season + season × body weight (Appendix Table 2.1).

No genetic terms were significantly associated with flea infection in the cross-sectional dataset. Individuals heterozygous at the *Il2* locus were somewhat less likely to be infected with fleas, but this association was marginally non-significant (β = -0.484 ± 0.273 s.e., p = 0.07).

The non-genetic minimal model examining flea infection in the longitudinal dataset was defined as: flea infection ~ site + season + sex +body weight + site × sex + *site*session* + *individual* (Appendix Table 2.3), observations = 512, groups (i.e. number of individuals) = 324. Two hundred and six voles (59.0%) were recorded as infected with fleas on at least one capture.

A significant association between variation at the *Il1b* locus and flea infection probability was observed under both additive (ΔAIC = 2.2) and heterozygosity models (ΔAIC = 2.1); heterozygotes had a decreased chance of infection, whereas individuals with the GGC haplotype were more likely to be infected (Table 4.2).

Flea burden

Total flea intensity per vole in the cross-sectional study ranged from 0 to 13 and had an overdispersed distribution, with 151 (49.2%) observations being zero, and only ten voles (3.3%) carrying >5 fleas. The minimal, non-genetic model was fitted with quasi-Poisson errors and was defined as: flea burden ~ site + season + weight. There was a strong seasonal effect on flea abundance with summer and autumn associated with the highest numbers. There were higher flea numbers at the BLB site and weight was positively correlated with flea burden (Appendix Table 2.2).

Genetic associations with flea burden were found under heterozygosity models for the genes *Il2* ($p = 0.048$) and *Slc11a1* ($p = 0.046$). Heterozygosity at both these loci was associated with a decreased number of fleas (Table 4.2).

4.4.4 Ticks

Tick infection

In the cross sectional study, 81 voles (26.4%) harboured at least one tick, with an extremely skewed distribution and a range of 0 to 108 ticks per animal. The non-genetic base model for probability of infection was: tick infection ~ season + body weight + eye lens weight + sex + season × eye lens weight + season × sex. Larger animals were more likely to be infected with ticks, while older animals were more likely to be infected in summer (Appendix Table 2.1).

A significant association between probability of tick infection and the *Il1b* locus was detected under a heterozygosity model ($p = 0.021$); individuals heterozygous at the *Il1b* locus were more likely than homozygotes to harbour a tick infection (Table 4.2). There was a marginally non-significant association between the *Tnf* locus and tick infection, with the *Tnf* T haplotype increasing the probability of infection compared to the C haplotype ($\beta = 0.432 \pm 0.236$ s.e., $p = 0.068$).

For the longitudinal dataset, 86 animals (24.6%) were positive for tick infection on at least one occasion. The minimal non genetic model was defined as: tick infection ~ body weight + *site*session* + *individual*, observations = 538, groups = 339.

Larger or older animals were more likely to be infected by ticks (Appendix Table 2.3).

An association between *Il2* and tick infection was observed under an additive model (ΔAIC = 2.3); those field voles with the rarest TC haplotype were less likely to be infected than those with the most common AC haplotype (Table 4.2).

Tick burden

The most parsimonious non-genetic model for tick burden in the cross-sectional study was defined as: tick burden ~ season + weight + eye lens weight + season × weight + season × eye lens weight + body weight × eye lens weight. There were significantly fewer ticks in summer and larger animals harboured greater numbers (Appendix Table 2.2).

Several significant associations between genetic loci and tick burden were found. Addition of the *Il2* term had a significant effect on the model of tick intensity under the heterozygosity model ($p = 0.038$), with heterozygotes at this locus associated with an increased tick burden (Table 4.2), and a marginally non-significant effect under the additive model ($p = 0.068$) with the rarest haplotype – TC – also associated with increased tick number (Table 4.2). Variation at the *Slc11a1* gene was associated with tick burden under a model of heterozygosity ($p = 0.015$) with heterozygous individuals associated with greater numbers of ticks (Table 4.2).

4.4.5 *Babesia microti*

B. microti infection was a binary presence/absence response, with the minimal non-genetic model for the cross sectional data given as: *B. microti* infection ~ season + body weight. Sixty-eight (22.2%) of the voles tested for *Babesia* were positive. There was a significantly lower probability of infection in summer 2008, while weight was positively correlated with the chance of *Babesia* infection (Appendix Table 2.1). There were no genetic terms significantly associated with *B. microti* in the cross-sectional dataset.

In the longitudinal study, 26.4% of voles were positive for *B. microti* on at least one occasion (*n* = 92). The model selected after deletion of non-significant terms was defined as: *Babesia microti* infection ~ site + season + sex + body weight + recapture + site × season + site × sex + *site*session*. *Individual* was not included as a random effect due to problems with model convergence. Larger animals and those captured previously were more likely to be infected (Appendix Table 2.3).

Strong genetic associations with *B. microti* infection were found with *Il2* under a heterozygosity model (ΔAIC = 7.2) and *Il12b* under an additive model (ΔAIC = 6.8). *Il2* heterozygotes were more likely to be infected with *Babesia* while voles with a CC *Il12b* haplotype had an increased chance of infection compared to the most common haplotype (Table 4.2).

4.4.6 *Bartonella*

Bartonella infection was recorded as presence/absence data. In total, 50 of the animals (16.3%) tested positive in the cross-sectional study. The non-genetic base model was: *Bartonella* infection ~ site + season + eye lens weight + season × eye lens weight. Voles caught at the SQC site were more likely to be infected with *Bartonella* and a significant interaction between season and eye lens weight was observed (Appendix Table 2.1). No genetic terms were associated with infection by *Bartonella* in the cross-sectional dataset.

The minimal model for the longitudinal data was *Bartonella* infection ~ season + sex + body weight + sex × weight + *site*session* + *individual*. Larger males and voles caught in summer were more likely to be infected. Two hundred of the captured voles (57.3%) were positive on at least one occasion for *Bartonella*.

Addition of *Il1b* to the model had a significant effect (ΔAIC = 4.6), as did *Il2* (ΔAIC = 5.9), both under heterozygosity models. For both *Il1b* and *Il2*, heterozygous individuals had a decreased chance of infection with *Bartonella* (Table 4.2).

4.4.7 Spatiotemporal variation in parasite prevalence

Temporal variation in prevalence was observed for every parasite examined in this study (Fig. 4.1). The prevalence of most parasites was lowest in winter and peaked in the summer, although *Babesia microti* exhibited a sharp dip in summer 2008. Frequency of parasite infection also varied spatially, with marked differences in prevalence between the two sites in some cases (Table 4.3). Most conspicuous were the data for nematodes, which showed that rates of host infection at BLB site were almost double that of SQC.

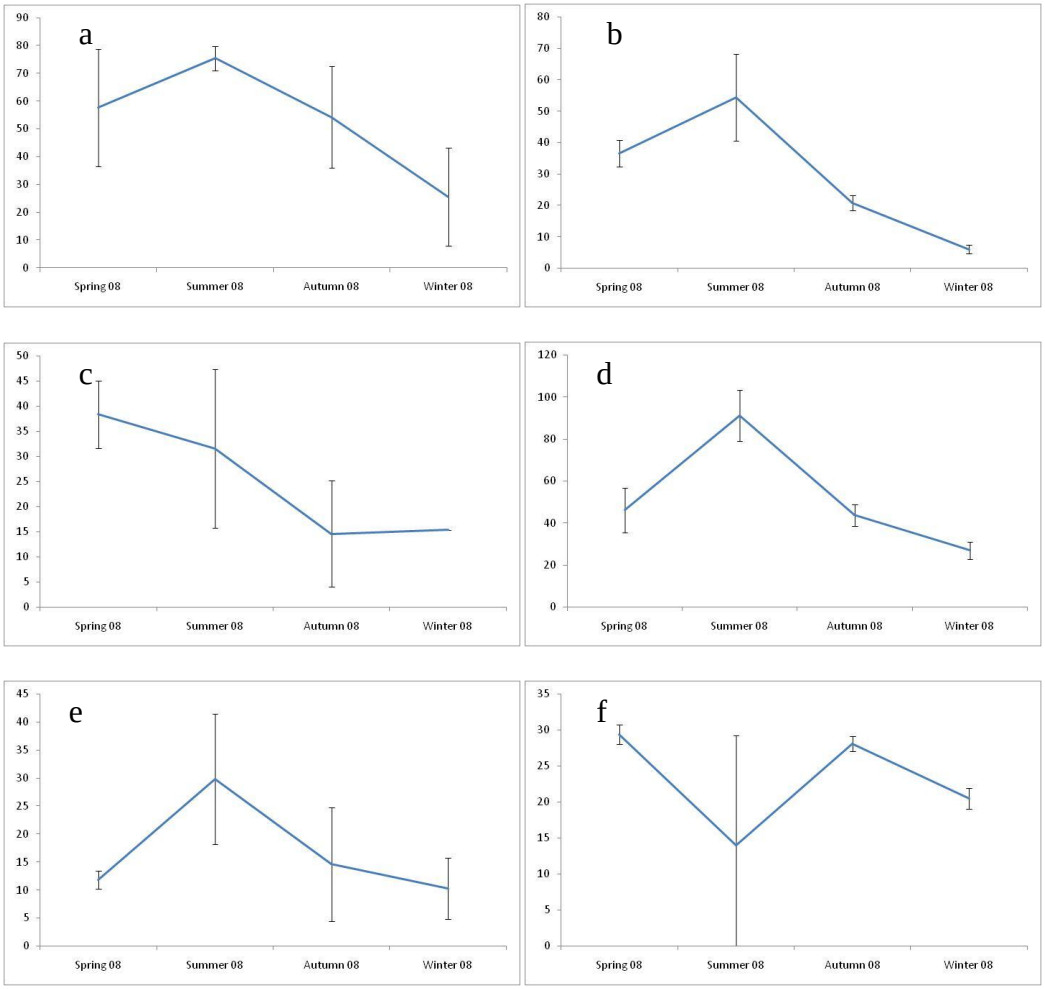

Figure 4.1 Seasonal variation in the percentage of animals positive for infection by (a) fleas, (b) ticks, (c) nematodes, (d) cestodes, (e) *Bartonella* spp and (f) *Babesia microti.* Data are average of combined BLB and SQC sites. Error bars represent standard deviation of the mean.

Table 4.3 Prevalence of various parasites at two sites in Kielder Forest

		Prevalence %		
Parasite	**Dataset**	**Total**	**BLB**	**SQC**
Bartonella spp.	CS	16.3	10.5	21.9
	Long.	57.3	52.9	61.8
Babesia microti	CS	22.2	25	19.3
	Long.	26.5	21.5	30.4
Fleas	CS	50.8	53.3	48.4
	Long.	59.0	55.1	62.3
Ticks	CS	26.4	27.6	25.2
	Long.	24.6	20.3	28.3
Nematodes	CS	21.5	27.6	15.5
Cestodes	CS	48.5	48.0	49.0

4.5 DISCUSSION

A better understanding of the genetic basis of susceptibility to infectious disease in natural populations is of great importance in conservation, ecology and evolutionary biology. Thus far, the majority of studies assessing how variation at immune-function loci affects infectious disease susceptibility in wild species have concentrated solely on genes of the MHC. However, there are a great number of other genes, including those encoding cytokines, which are critical to the immune response; genetic polymorphism within these genes that results in an alteration in structure or expression levels can have pathological consequences to the host, including the potential for increased risk of infection (Smith and Humphries 2008). Here, for the first time, I describe genetic associations between several non-MHC immune genes and various parasites within a natural population of the field vole, *Microtus agrestis*. The loci most consistently associated with variation in susceptibility to parasites were genes at which analyses of previous chapters have identified patterns of diversity consistent with balancing selection, namely *Il1b* and *Il2*. Host-parasite interactions therefore provide a plausible mechanism behind the maintenance of diversity observed at these genes in a natural population.

4.5.1 Genetic associations with parasite resistance

The main finding of this chapter is that both *Il1b* and *Il2* repeatedly exhibit associations with resistance to pathogens. Polymorphism within both genes was consistently associated with variation in susceptibility to a range of naturally-occurring parasites in Kielder Forest; the two loci between them accounted for 63% of the total number of significant associations with the remaining 37% split between five other immune genes. Genetic variation at *Il1b* was associated with varying susceptibility to fleas, ticks, and *Bartonella*, while associations between *Il2* and numerous pathogens were observed: polymorphism within *Il2* was associated with variation in the probability of infection by nematodes, cestodes, ticks, *Babesia microti* and *Bartonella*, as well as variation in total burdens of ticks and fleas.

Earlier chapters have provided strong evidence suggesting that balancing selection is acting to maintain diversity at both *Il1b* and *Il2*. In the previous chapter I hypothesised that the most likely form of balancing selection acting upon these genes is fluctuating selection for different alleles, caused by spatiotemporal variation in parasite abundance, as allele frequencies differ between neighbouring subpopulations at these loci. The consistency of associations with a range of pathogenic organisms that do exhibit spatial and temporal variation in abundance supports the hypothesis that parasites are likely to be the selective force driving the maintenance of genetic diversity at the field vole *I1lb* and *Il2* genes.

Previous chapters have also accumulated strong evidence for the patterns of diversity at *Tnf* having been maintained by balancing selection. However, allele frequencies did not differ significantly between subpopulations as did those at *Il1b* and *Il2*, which suggests that fluctuating selection is probably not the form of balancing selection acting on this particular gene. *Tnf* was strongly associated with cestode resistance, with the T haplotype decreasing the chance of acquiring an infection with cestodes, but with no other parasites. However, it should be noted that Kielder Forest field voles are infected with numerous other endemic parasites including cowpox virus (Cavanagh *et al.* 2004; Burthe *et al.* 2006; Burthe *et al.* 2008b), tuberculosis (*Mycobacterium microti*) (Cavanagh *et al.* 2002; Cavanagh *et*

al. 2004; Burthe *et al.* 2008a), trypanosomes (Smith *et al.* 2005) and a variety of other ectoparasites including lice and mites. Whether infection by cestodes exerts a strong enough selective pressure to be the force underlying maintenance of diversity at *Tnf*, or whether other parasites or environmental factors not measured in this study are driving the process, is as yet unclear.

Of the other genes, *Tlr4* has shown signs of positive directional selection in previous chapters, including the observation that the derived allele at the non-synonymous Tlr4 1663 A/G SNP has been driven almost to fixation, which may imply functional relevance at this or a linked SNP. Here, no associations with pathogen susceptibility and *Tlr4* variation were observed, which would imply that parasites, or at least those measured in this study, are not the cause of this putative positive selection. The Il12b 704 C/T SNP had previously been predicted as possibly deleterious to protein function by the two programs PolyPhen and SIFT (Chapter 2). As the T allele was found only within the BLB site, and even here at very low frequencies, this may be indicative of negative selection preventing the accumulation of this allele. The results of this chapter provide some support for the functional relevance of genetic diversity within *Il12b*, as variation at this locus was strongly associated with susceptibility to nematodes and *Babesia microti*, and also for negative selection as, for both parasite response variables, the rarest haplotypes were associated with increased risk of infection. Genetic associations between several other genes and parasite susceptibility were also observed, although further investigation would need to be done to ascertain fully whether these are true associations or artefacts of multiple testing.

Multiple testing is somewhat of a contentious issue in association studies, in that the more tests one carries out the more likely one is to generate a false positive (a type I error); correcting for the number of tests is therefore common. A reasonably large number of tests of association were carried out in this study. However, a Bonferroni correction or similar would likely be too conservative and lead to the rejection of truly significant associations (a type II error). Instead I have relied on an approach of accumulating consistent evidence of functionality; *Il1b* and *Il2* have

shown strong evidence for balancing selection over several tests of neutrality in previous chapters, and after hypothesising that parasites were the main selective force driving this, have here been consistently associated with variation in resistance to a range of parasites. It is therefore likely that the polymorphism observed within these two genes is functionally relevant and mediates phenotypic variation between individuals in parasite resistance. Genetic associations with infection by parasites were also observed at other loci, but less consistently.

4.5.2 Immunology of host-parasite interactions

Although no previous studies have examined associations between *Il1b* and *Il2* and infectious disease in a wild species, their crucial roles in immunity means it is not difficult to imagine that polymorphism within these genes may have phenotypic consequences. IL-1β is a pleiotropic pro-inflammatory cytokine that has a central and multifunctional role in host immunity, including mediation of inflammatory responses, stimulation of macrophages and activation of T-cells through enhancing the production and expression of IL-2 and its receptor (Borish and Steinke 2003). IL-2 is secreted by T-cells in response to accessory signals provided by IL-1 and other molecules, and acts in concert with the IL-2 receptor (IL-2R) to induce clonal expansion of antigen specific T-cells. Other functions of IL-2 include activation of B-cells, natural killer (NK) cells and macrophages (Borish and Steinke 2003).

Infection by both *Bartonella* and *Babesia* has previously been shown to have tangible effects on host production of IL-1β and IL-2. Internalization of *Bartonella henselae* by macrophages has been shown to induce high levels of IL-1β release in both murine and human cell lines (Musso *et al.* 2001; Resto-Ruiz *et al.* 2002) while *Bartonella quintata* lipopolysaccharide (LPS) inhibited the production of IL-1β by *Escherichia coli* LPS-stimulated human monocytes through antagonism of TLR4 (Popa *et al.* 2007), suggesting that *Bartonella* may also be able to down-regulate host immunity. In addition, circulating levels of IL-2 are higher in patients infected with *Bartonella henselae*. Likewise, infection by *Babesia* spp. is associated with increased expression of IL-1β and IL-2 (Shoda *et al.* 2000; Iseki *et al.* 2008). In this study *Il2* was also significantly associated with infection by nematodes and

cestodes. Previous work has shown that in mitogen-stimulated splenocyte cultures, field vole *Il2* expression is strongly associated with mediators of the Th1 response, which is involved primarily in immunity against microparasites (Jackson *et al.* in review). However, IL-2 also plays a central role in the differentiation of T-helper cells into the Th2 subtype, which is critical for an effective response to helminth infection (Cote-Sierra *et al.* 2004; Paul and Zhu 2010). It is likely therefore that IL-2 is important in supporting many branches of the adaptive immune response in field voles; the association between *Il2* polymorphism and helminth infection may arise from differential expression altering the balance between Th1 and Th2 responses in the host.

Variation within the cytokine genes *Il1b* and *Il2* was also significantly associated with susceptibility to ectoparasites. It has previously been shown that blood-feeding ectoparasites can modulate host cytokine production (Schoeler *et al.* 1999; Wikel 1999; Jackson *et al.* 2009) and have a demonstrable effect on both host fitness and the scope of the immune response. For example, in the common vole *Microtus arvalis*, flea infestation can induce anaemia, affect growth and body size and depress the host immune system (Devevey *et al.* 2008), leading ultimately to a decreased life span for infected individuals (Devevey and Christe 2009). Furthermore, variation at MHC loci in the water vole *Arvicola terrestris*, is associated with burdens of several ectoparasites including fleas and ticks (Oliver *et al.* 2009), demonstrating that immune gene polymorphism can affect levels of ectoparasitism in natural vole populations. The immunosuppression of host immune responses mediated by parasitic arthropods may provide a more hospitable environment for pathogens for which they act as a vector (Wikel 1999). The endemic ectoparasites inhabiting Kielder Forest act as vectors for numerous parasites capable of infecting field voles; whether the genetic associations with fleas and ticks observed here are due to direct interaction between the host immune system and the ectoparasites themselves, or through an immune response to pathogens which they carry, is not yet known.

4.5.3 Possible mechanisms of parasite-driven selection at *Il1b* and *Il2*

The most compelling evidence for phenotypic variation resulting from polymorphism in field vole immune genes is for *Il1b* and *Il2*, where genetic diversity at both loci appears to have been maintained by parasite-driven selection. Due to the critical role of cytokines in the immune response, one may expect purifying selection to be the main force shaping their patterns of genetic diversity, acting as it does to prevent the accumulation of deleterious mutations and to maintain function. Why then is selection acting to maintain genetic diversity at *Il1b* and *Il2*? One possibility is that the maintenance of balanced polymorphism within these two genes may be related to the multifunctional nature of the cytokines they encode. Unlike TLRs and similar molecules, cytokines do not interact directly with pathogens and are therefore unlikely to undergo selective sweeps in response to a molecular change on a specific pathogen. Cytokines instead function as messenger molecules and are critical mediators of every aspect of the vertebrate immune response. They interact with a wide range of molecules including other cytokines and immune receptors and their actions include initiating and regulating both pro- and anti- inflammatory effects, and humoral and cell-mediated immune responses. Alongside their direct immunological roles cytokines have a wide-range of non-immune physiological effects including an influence on thermoregulation, appetite and fatigue (Corwin 2000). It is therefore likely that an alteration in the action of cytokines will have a more significant effect on host pathology than simply its impact on infectious disease susceptibility (Downing *et al.* 2010). Mediation of the inflammatory response is essential in maintaining the balance between health and disease and as such the inflammatory response is finely-tuned and tightly regulated (Ferrer-Admetlla *et al.* 2008). An effective immune response is essential for the clearance of pathogenic organisms, but too strong a response can cause a variety of adverse effects related to auto-immunity and the complications of infectious disease. For example, IL-1β production is known to induce a spectrum of symptoms associated with being unwell, including lethargy, sleep and fever, through interaction with the central nervous system (Borish and Steinke 2003). Both IL-1β and IL-2 are broadly immunostimulatory in function and have been linked with pathogenesis in several autoimmune and inflammatory diseases (O'Shea

et al. 2002; Kunz and Ibrahim 2009). Increased expression of one or both of these cytokines could then theoretically be advantageous in an environment containing a high number of parasites if an increased inflammatory response resulted in an increased likelihood of surviving infection; conversely, in the absence of particular infectious organisms, a genotype that up-regulates inflammatory responses may be deleterious, resulting in pathology to the host unrelated to infection. The high levels of genetic diversity maintained within *Il1b* and *Il2* may then regulate a balance between normal physiological functioning and an ability to deal effectively with infectious disease, in a habitat exhibiting spatiotemporal variation in parasite abundance.

It is perhaps unlikely that a single parasite will exert a high enough selective pressure to shape the patterns of diversity observed here in a subset of field vole cytokine genes. Co-infection by several parasites is the norm for natural populations, and interactions between several pathogens has previously been shown to influence infection risk in field voles (Telfer *et al.* 2008; Telfer *et al.* in press). Furthermore, a recent study has shown that naturally-occurring parasites are capable of exerting immunosuppressive effects on wild rodent populations (Jackson *et al.* 2009), which may enable other parasites to better survive the host immune response. The mechanism by which natural selection has shaped the genetic diversity at *Il1b* and *Il2*, and possibly other genes, is therefore likely to have arisen from a complicated web of interactions between host genotype, a spectrum of parasites and a heterogeneous environment.

Exactly how variation at immune gene loci affects susceptibility to parasites is beyond the scope of this chapter but will be investigated in the next. As mentioned earlier, cytokines do not interact directly with parasites and therefore cytokine coding sequence is likely to be under strong purifying selection to maintain molecular function. Variation in cytokine activity may therefore be more likely to be due to polymorphism having an effect on gene expression and the subsequent immune response; the next chapter will therefore examine the relationship between immunogenetic diversity and immune function.

4.6 CONCLUSION

Non-MHC immune genes have been underrepresented in studies of infectious disease susceptibility in natural populations, despite their clear importance in the immune response. In this chapter I have presented evidence that variation at several field vole immune genes, in particular *Il1b* and *Il2*, is functional and associated with variation in susceptibility to a range of parasites. Coupled with the support for balancing selection at these loci, these results provide compelling evidence that host-parasite interactions are driving the maintenance of genetic diversity in some field vole immune genes.

Chapter 5

Immunogenetic diversity and variation in immune phenotype

5.1 INTRODUCTION

A host's general ability to resist infection, or immunocompetence, is often seen as a key component of fitness (Grenfell *et al.* 2002). However, immunocompetence is unlikely to be a simple quantity equally applicable to all types of pathogen, and instead the type of immune response elicited rather than overall immune responsiveness is more likely to determine host resistance (Bradley and Jackson 2008). The vertebrate immune system is multifaceted and complex, with a particular immune response resulting from a complicated interplay between the innate, adaptive and regulatory arms of the immune system. Resistance to one pathogen resulting from phenotypic commitment to a particular functional immune response may induce susceptibility to another parasite, controlled by different effector mechanisms (Bradley and Jackson 2008). As such, the ability to measure and account for different aspects of the immune response (innate, adaptive, regulatory) in natural populations, rather than rely on a simplified measure of immunocompetence, may therefore help us to better define the immune status of hosts and increase our understanding of how this influences life history traits, host-parasite interactions, co-infection dynamics and disease transmission (Graham *et al.* 2007; Bradley and Jackson 2008; Jackson *et al.* in review).

Analysis of cytokine profiles provides a useful way to simplify the complexity of the immune response without sacrificing realism (Graham *et al.* 2007). These immune signalling molecules mediate all aspects of the immune response and are critical in the polarisation and amplification of the immune system. As a result, cytokines help to determine which effector mechanisms are employed in response to a given immune insult (Graham *et al.* 2007). The balance between Th1- and Th2-promoting cytokines is a major factor in the outcome of the host response to

infection and any genetic variants that influence this balance and the resulting immune phenotype are likely therefore to be a major factor in determining host susceptibility (Abbas *et al.* 1996; Charbonnel *et al.* 2006).

The previous chapter demonstrated that genetic variation in a subset of cytokine genes of the field vole is associated with resistance to a range of parasites. Given the central immunological role of cytokines, one may reasonably assume that these observed effects are mediated by an individual's immune phenotype. This chapter will examine the relationship between genetic diversity and immune function, measured through analysis of mRNA expression of several cytokines and transcription factors, as a possible explanatory mechanism for the variation in parasite resistance observed in the field vole.

In the overwhelming majority of non-model species, pre-existing antibody reagents are lacking and the development of novel, conventional antibody-based methods for detecting proteins of interest are often prohibitive in terms of cost, time or resources (Oko *et al.* 2006). Measurement of mRNA expression through quantitative real-time reverse-transcription PCR (Q-PCR) offers a reliable alternative to the direct measurement of proteins for studies involving natural populations. An important caveat of all gene expression studies is that mRNA expression does not reflect the full range of regulatory process involved in protein production, as many genes are regulated during translation; processes such as protein modification which occur post-transcription and may affect the amount of active protein are not taken into account when measuring mRNA expression. However, cytokines and transcription factors are often found at such low levels that mRNA quantification represents the only technique sensitive enough to detect their expression *in vivo* (Huggett *et al.* 2005).

In humans, polymorphisms within numerous cytokine genes have previously been associated with differential patterns of expression (Smith and Humphries 2008) which in turn can lead to indirect effects on downstream events in the immune response (Kelso 1998). If polymorphism within field vole cytokines has a similar

effect, then variation in immune phenotype may be the mechanism underlying the observed variance in pathogen susceptibility. As such, I will test the hypothesis that those genes consistently associated with parasite resistance in the previous chapter may also be associated here with variation in immune gene expression.

5.2 OUTLINE

The aim of this chapter is to examine the association between genetic diversity and variation in immune phenotype in a natural field vole population. As polymorphism within a gene can potentially affect the transcription of either itself or other genes, associations with and between a range of immunity genes are examined. Previous chapters have shown that genetic diversity at *Il1b* and *Il2* appears to have been shaped by parasite-driven selection, and that there is some evidence that variation within *Il12b* is linked to parasite resistance; if the physiological mechanism underlying these observations is variation in immune function, then these loci may be associated with variation in gene expression.

5.3 MATERIALS AND METHODS

Animal trapping and immunological measurements were carried out as part of a separate study investigating immunodynamics in a natural field vole population (Jackson *et al.* in review) and as such the methodology for these aspects of the study will be described relatively briefly here.

5.3.1 Sampling site and procedure

Naturally-occurring Kielder Forest field voles were trapped and processed as part of the cross-sectional study described in the previous chapter and more fully in Jackson *et al.* (in review).

5.3.2 Gene expression assays

Relative mRNA expression of several cytokines and immunity-related transcription factors was measured from cultured leucocytes isolated from individual field vole

spleens (full description of cell culture methods can be seen in Jackson *et al*. (in review)). Splenocytes were cultured for 24 or 96 hours depending on the assay (see below). Messenger RNA accumulations were measured from both unstimulated cells, reflecting constitutive levels of mRNA expression, and from cells cultured with various immunostimulatory molecules, under the assumption that observed *ex vivo* responses reflect the potential for that response to occur *in vivo* (Jackson *et al.* in review).

Constitutive mRNA accumulations of the following genes were measured at 24 hours: *Il1b*, *Irf5* (which codes for interferon regulatory factor 5, a transcription factor which regulates the production of type I interferons and inflammatory cytokines (Paun *et al.* 2008)), *Tgfb1* and *Il10*. Alongside these measurements of mRNA accumulations in unstimulated cells, expression was also measured from splenocyte cultures stimulated with either the TLR2 agonist HKLM (heat-killed *Listeria monocytogenes*) or the synthetic TLR7 agonist, imiquimod (InvivoGen, San Diego, USA) (Table 5.1). The production of *Irf5* mRNA in the TLR2-stimulated culture and *Il1b* in TLR7-stimulated cells did not demonstrate a positive response to the stimuli after analysis of 100 individuals and so these measurements were ceased (Jackson *et al.* in review).

For the 96-hour assays, measurements were taken of mRNA accumulations of *Ifng*, *Tbx21* (also known as *Tbet*; codes for the transcription factor T-box 21, which is specific to the Th1 cell subset (Mullen *et al.* 2001)), *Gata3* (encodes GATA binding protein 3, a transcription factor associated with the Th2 cell subset (Zhu *et al.* 2006)), *Foxp3* (another transcription factor, forkhead box P3, associated with regulatory T-cells (Hori *et al.* 2003), *Tgfb1* and *Il10*. Measurements of mRNA accumulations were taken from both unstimulated splenocyte cultures and cultures stimulated with the mitogen phytohaemagglutinin (PHA), which preferentially induces proliferation of activated $CD4^+$ helper T-cells (O'Donovan *et al.* 1995) (Table 5.1).

Table 5.1 Summary of splenocyte culture assays and mRNA measurements taken. Unstim., unstimulated cell culture; HKLM, culture stimulated with the TLR2 agonist HKLM; Imiqui, culture stimulated with the TLR7 agonist imiquimod; PHA, culture stimulated with the mitogen PHA.

	24 h splenocyte assay			96 h splenocyte assay	
Gene	**Unstim.**	**HKLM**	**Imiqui.**	**Unstim.**	**PHA**
Il1b	✓	✓	-	-	-
Irf5	✓	-	✓	-	-
Il10	✓	✓	✓	✓	✓
Tgfb1	✓	✓	-	✓	✓
Ifng	-	-	-	✓	✓
Tbx21	-	-	-	✓	✓
Gata3	-	-	-	✓	✓
Foxp3	-	-	-	✓	✓

For each assay described above, relative mRNA accumulations in splenocyte cultures were measured by two-step reverse-transcription quantitative real-time PCR (Q-PCR) (see Jackson *et al.* in review for full details). Briefly, target genes were first normalised to the endogenous control gene *Ywhaz* (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide). Immune gene mRNA expression was then measured and analysed as normalised expression relative to a reference, pooled cDNA sample (the ΔΔCt method, Livak and Schmittgen 2001). This method allows one to compare relative, normalised gene expression levels between animals (Jackson *et al.* in review).

5.3.3 Genotyping

Genomic DNA was extracted from the livers of freshly dissected field voles using the DNeasy Blood and Tissue Kit (QIAGEN). Eighteen SNPs within eight immune genes were genotyped by KBioscience (for full details see Chapter 3, section 3.3.3).

5.3.4 Statistical analysis

There was a high level of redundancy in the multivariable dataset; in almost all cases where multiple measurements were taken on expression of the same gene (for example, innate, unstimulated expression and TLR-stimulated expression levels)

these measurements were significantly correlated. Therefore, two techniques were used to group variables together to minimize the number of tests performed: for genes where two, correlated measurements of gene expression were taken (*Foxp3*, *Ifng*, *Il1b*, *Il2*, *Irf5* and *Tbx21*) both measurements were standardised by, for each value, subtracting the mean and dividing by the standard deviation. These two standardised measurements were then summed to give a single response variable for each gene. Where more than two measurements were taken for a particular gene (*Il10* and *Tgfb*), principal components analysis (PCA) was used to reduce these multiple, correlated measurements into a single variable derived as the individual scores for a dominant first principal component (PC) (Jackson et al. 2009). Within each derived, reduced variable, component loadings were of a similar magnitude and of the same sign, indicating a pattern of positive covariation. All variables were log-transformed (Log_{10} (x + 1)) prior to standardization or PCA. *Gata3* was an exception in that two measurements were taken which did not correlate and therefore each *Gata3* response variable was analysed separately.

Statistical modelling was performed using R version 2.10 (R Development Core Team 2009). The primary aim of the analyses was to examine the relationship between immunogenetic variation and gene expression, whilst controlling for confounding non-genetic factors. The underlying method for each analysis was then the same as that of the previous chapter examining genetic associations with parasite resistance; first, a minimal model containing only non-genetic terms was constructed to account for confounding factors which may influence immune function and gene expression. To this minimal, non-genetic model each genetic term was added in turn to assess their significance in explaining the variance in expression of a given gene.

Analyses of the immune response measures were broadly the same (aside from *Gata3*, described below). Linear models (LM) assuming normal errors were used to construct non-genetic models for each of the immune response variables, either the summed standardised scores or the scores for the dominant first PC. As for the parasite data in the previous chapter, a maximal non-genetic model was constructed

containing site, season, sex, body weight, eye lens weight and their two-way interactions. The significance of each non-genetic variable was again assessed by comparing the deviance explained by the full model to that of a model which did not contain the term. The full non-genetic model was reduced by removing each non-significant term, beginning with interactions and terms that had the highest *p*-value; terms were deletion tested such that those that did not cause a significant increase in deviance when removed were left out of the model (see Chapter 4, section 4.3.5 for full details). Following deletion testing only significant terms were left in the resulting non-genetic minimal model.

As the two *Gata3* measurements were not correlated they were not grouped and were analysed separately. Unstimulated 96 h *Gata3* expression could not be normalised by any conventional transformation and was therefore analysed as a binary response variable using a GLM and a binomial error structure. The binary response corresponded to presence/absence, where presence was recorded as a detectable change in expression compared to the reference cDNA sample. The mitogen-stimulated 96 h *Gata3* data was normalised by performing a Box-Cox power transformation (Box and Cox 1964) within the MASS library (Venables and Ripley 2002). The transformed data were analysed using a linear model as described above. Both *Gata3* models were deletion tested as before until only significant terms remained in the resulting non-genetic minimal model.

Principal components analysis was also used to derive single variables for both general pro-inflammatory and anti-inflammatory responses, termed PC^{I} and PC^{AI}, respectively (Jackson *et al.* in review). Here, each of the two dominant first principal components incorporated correlated measurements from several genes in order to examine the relationship between genetics and a more general immune response, rather than expression of a single gene.

After construction of the non-genetic models, significance of genetic terms was calculated by addition of these terms to the minimal non-genetic model. Analysis of deviance tests were then used to assess whether the model including genetic

terms was a significantly better fit to the data than the model without. Genetic terms were fitted to the minimal models under either a heterozygosity model, where, for each gene, values of heterozygotes were compared to homozygotes, or an additive model. The additive model compared the most common allele (haplotype) at a particular locus to all other alleles; therefore, for a specific allele at a particular locus the value of the heterozygote class lies between the two homozygote classes (that is, given the comparison of an allele A_i against the most common allele, termed A_0, the genotypes A_0A_0, A_0A_i and A_iA_i have the genotype values 0, α_i and $2\alpha_i$, respectively, where α_i represents the additive effect of the A_i allele) (Paterson *et al.* 1998).

5.4 RESULTS

Splenocytes from 307 captured field voles (152 from the BLB site and 155 from SQC) were cultured and subjected to the various immunoassays described above. As the primary focus of this study was to identify genetic associations with immune gene expression and not to ascertain the importance of environmental and other non-genetic factors on gene expression levels, non-genetic factors will be only discussed briefly here. The purpose of the minimal non-genetic models in this study was to account for confounding environmental and intrinsic factors prior to the addition of genetic terms as explanatory variables. Table 5.2 summarises the retained non-genetic terms significantly associated with expression of the various field vole immune genes, while a fuller summary of the models can be found in Appendix Tables 2.5 and 2.6. For a more detailed discussion on the relationship between non-genetic factors and immunodynamics in this population of field voles see Jackson *et al.* (in review).

5.4.1 *Foxp3* expression

Two measurements of *FoxP3* expression were taken, unstimulated expression and mitogen-stimulated expression, both recorded at 96 h. As these two measurements were correlated ($r = 0.215$, $p = 0.005$) their standardised scores were combined and

reduced to a single response variable (see methods). The minimal, non-genetic model for *Foxp3* was defined as: *Foxp3* ~ site + season + body weight + eye lens weight + sex + site × body weight + site × eye lens weight + site × sex (Table 5.2). Expression of *Foxp3* was lowest in spring 2008 and peaked in autumn. Males and older animals (measured using eye lens weight as a proxy) at the SQC site had increased levels of *Foxp3* expression, while body weight was negatively correlated with expression at SQC (Appendix Table 2.5). No significant genetic associations with *Foxp3* expression were observed.

Table 5.2 Non-genetic terms significantly associated with immune gene expression. *P*-value not shown for main effect if the covariate is included in an interaction. Deletion testing led to the removal of all non-genetic terms in the models of *Il10* and *Irf5* expression.

Response	**Terms**	**d.f.**	***F*-value**	***p*-value**
	Linear models			
Foxp3	Season	4	3.16	0.015
	Site × body weight	1	5.49	0.020
	Site × eye lens weight	1	6.43	0.012
	Site × sex	1	5.06	0.026
Gata3, 96 h PHA[a]	Season × body weight	4	2.70	0.032
Ifng	Site	1	5.46	0.020
	Season	4	15.96	<0.001
	Body weight	1	3.94	0.049
Il1b	Season	4	11.56	<0.001
	Body weight × eye lens weight	1	6.25	0.013
Il2	Season	4	3.88	0.005
Il10	-	-	-	-
Irf5	-	-	-	-
Tbx21	Season	4	17.81	<0.001
	Sex	1	4.00	0.047
Tgfb1	Season × sex	4	2.51	0.045
PC^{I}	Season	4	14.61	<0.001
PC^{AI}	Season × sex	4	2.63	0.038
	Body weight × eye lens weight	1	6.05	0.015
	Generalised linear model			
		d.f.	**LRT**	***p*-value**
Gata3, 96 h unstim.[b]	Season	4	11.17	0.025
	Eye lens weight × sex	1	4.98	0.026

[a] *Gata3* expression measured at 96 h from splenocyte cultures stimulated with the mitogen PHA

[b] *Gata3* expression measured at 96 h from unstimulated splenocyte cultures

5.4.2 *Ifng* expression

Ifng expression was modelled as a single response variable formed by combining the standardised scores of unstimulated and mitogen-stimulated expression at 96 h. The two measurements were significantly correlated ($r = 0.482$, $p < 0.001$). The minimal model for *Ifng* expression was defined as: *Ifng* ~ season + site + weight (Table 5.2). Peak expression was observed in winter 2008 and spring 2009, while heavier animals and those captured at the SQC site exhibited increased expression levels (Appendix Table 2.5).

No significant genetic associations were observed with *Ifng* expression, although the *Il1b* locus had a marginally non-significant effect under the additive model ($p = 0.07$); voles which harboured the most common *Il1b* GGC haplotype exhibited the highest *Ifng* expression while, in comparison, the rarest GGT haplotype was associated with lower expression.

5.4.3 *Il1b* expression

The two measurements of *Il1b* expression recorded were unstimulated and TLR2-stimulated expression at 24 h. The two were correlated ($r = 0.384$, $p < 0.001$) and so were reduced to a single *Il1b* expression response variable. The minimal non-genetic model for *Il1b* expression was defined as: *IL1b* ~ season + body weight + eye lens weight + body weight × eye lens weight (Table 5.2). Expression of *Il1b* demonstrated a pronounced dip over spring and summer 2008 before increasing in winter and spring 2009 (Appendix Table 2.5).

There was a highly significant association between *Il1b* expression and the *Il1b* locus when fitted under an additive model ($p = 0.008$) but not a heterozygosity model; the *Il1b* GAT haplotype was associated with an increase in *Il1b* expression (a summary of coefficients for significant genetic terms can be found in Table 5.3). Addition of *Il12b* as a genetic term significantly improved the explanatory power of the model under both additive ($p = 0.003$) and heterozygosity ($p = 0.036$) models; the rarest *Il12b* GT haplotype is associated with higher *Il1b* expression than the

most common haplotype, GC, while *Il12b* heterozygotes exhibited increased levels of *Il1b* expression (Table 5.3). There was also a marginally non-significant association between *Il1b* expression and the *Il2* locus under the heterozygosity model ($p = 0.07$) where the trend was for *Il2* heterozygotes to express lower levels of *Il1b*.

Table 5.3 Genetic terms significantly associated with gene expression

Gene	Response	Comparisons	Coefficient (s.e.)	*t*-value	*p*-value
Il1b	*Il1b* expression	GAT vs. GGC	0.62 (0.20)	3.19	0.002
	Gata3 96 h expression	Heterozygote vs. homozygote	-1.08 (0.52)	-2.09	0.037
Il2	*Il10* expression	Heterozygote vs. homozygote	-0.41 (0.20)	-2.11	0.037
Il12b	*Il1b* expression	Heterozygote vs. homozygote	0.58 (0.28)	2.11	0.036
		GT vs. GC	1.56 (0.50)	3.14	0.002
	Il2 expression	Heterozygote vs. homozygote	-1.15 (0.37)	-3.14	0.002
		CC vs. GC	-1.17 (0.45)	-2.62	0.010
Tnf	*Gata3* 96 h expression	Heterozygote vs. homozygote	-1.02 (0.53)	-1.94	0.052

5.4.4 *Il2* expression

The *Il2* response variable was derived from the summed standardised scores of 96 h unstimulated expression and 96 h mitogen-stimulated expression, which were significantly positively associated ($r = 0.435$, $p < 0.001$). The minimal non genetic model explaining *Il2* expression was: *Il2* ~ season (Table 5.2), with higher *Il2* expression associated with winter 2008 and especially spring 2009 (Appendix Table 2.5).

The *Il12b* locus was significantly associated with *Il2* expression under both additive ($p = 0.007$) and heterozygosity ($p = 0.002$) models. Heterozygotes at the *Il12b* gene exhibited much lower *Il2* mRNA accumulations than homozygotes and, related to this, the rare CC and GT haplotypes displayed lower *Il2* expression compared to the most common GC haplotype (Table 5.3).

5.4.5 *Il10* expression

As several measurements of *Il10* expression were taken, PCA was used to reduce these multiple, correlated measurements into a single variable derived as the individual scores for a dominant first principal component of *Il10* expression ($Il10^{PC1}$) (see methods). The PCA of log-transformed *Il10* measurements produced a first principal component with an eigenvalue of 1.51 and that explained 38% of the variation in *Il10* expression. The variable loadings for $Il10^{PC1}$ were: 24 h unstimulated = 0.498; 24 h TLR2-stimulated = 0.553; 24 h TLR7-stimulated = 0.616; 96 h unstimulated = 0.059; 96 h mitogen-stimulated = 0.251. 96 h unstimulated expression was excluded from the analysis as the component loading (0.059) was small. The minimal, non-genetic model of *Il10* expression explained very little of the variation in *Il10* expression as deletion-testing allowed removal of all terms. The minimal model was therefore the null model, $Il10^{PC1} \sim 1$.

Genetic terms were therefore added to the null model. There was a significant association between *Il10* expression and the *Il2* locus under the heterozygosity model ($p = 0.037$). *Il2* heterozygotes presented lower *Il10* expression levels than homozygotes (Table 5.3).

5.4.6 *Irf5* expression

Two correlated measurements of *Irf5* expression (24 h unstimulated and 24 h TLR7-stimulated expression; $r = 0.395$, $p < 0.001$) were combined into a single response variable. No non-genetic terms were retained after deletion testing resulting again in the null model as the minimal model. Addition of genetic terms to this null model did not result in any associations with *Irf5* expression.

5.4.7 *Gata3* expression

There was no significant correlation between the two measurements of *Gata3* expression, 96 h unstimulated and 96 h mitogen-stimulated expression ($r = -0.107$, $p = 0.132$). As such, the two responses could not be combined and reduced to a single variable and were therefore analysed separately.

Unstimulated 96 h *Gata3* expression could not be normalised and was therefore analysed as a binary, presence/absence response variable using a GLM assuming binomial errors. The non-genetic minimal model was defined as: *Gata3* 96 h unstim. ~ season + eye lens weight + sex + eye lens weight × sex (Table 5.2). The least likely capture season to observe a detectable response was winter 2008 and, while males were less likely to exhibit an observable *Gata3* response than females generally, larger males were associated with an increased probability (Appendix Table 2.6). Genetic associations with unstimulated 96 h *Gata3* expression were observed for *Il1b* ($p = 0.040$) and *Tnf* ($p = 0.041$) under a model of heterozygosity. For both loci, heterozygous individuals were less likely to show a detectable response (Table 5.3.)

Mitogen-stimulated 96 h *Gata3* expression was normalised by using a Box-Cox procedure (Box and Cox 1964) to find the optimal power transformation ($\lambda = -2.16$). The normalised data was then analysed using a general linear model assuming normal errors. The minimal model was defined as *Gata3* 96 h PHA ~ season + body weight + season × body weight (Table 5.2) with heavier animals associated with increased expression in summer (Appendix Table 2.5). No genetic associations were observed.

5.4.8 *Tbx21* expression

Two correlated measurements of *Tbx21* expression were recorded, 96 h unstimulated and 96 h mitogen-stimulated expression ($r = 0.324$, $p < 0.001$) and grouped into a single response variable. The minimal model for *Tbx21* expression was defined as: *Tbx21* ~ site + season (Table 5.2). Animals from the SQC site displayed higher expression than those at BLB. Expression levels decreased from spring through summer and autumn 2008 before increasing in winter 2008 and peaking in spring 2009 (Appendix Table 2.5). There were no genetic associations with *Tbx21* expression.

5.4.9 *Tgfb1* expression

Principal components analysis was again used to group multiple *Tgfb1* expression variables into a single response variable, $Tgfb1^{PC1}$. The resulting dominant first PC had an eigenvalue 1.38 and explained 34% of the variation in *Tgfb1* expression. Component loadings for $Tgfb1^{PC1}$ were: 24 h unstimulated = 0.610; 24 h TLR2-stimulated = 0.527; 96 h unstimulated = 0.305; 96 h mitogen-stimulated = 0.508. The best base model was defined as: $Tgfb1^{PC1}$ ~ season + sex + season × sex (Table 5.2). Males exhibited significantly lower expression of *Tgfb1* in winter (Appendix Table 2.5). No significant genetic associations were found, but the *Il1b* locus was associated at a marginally non-significant level under the additive model (p = 0.056), with individuals carrying the rarest, GGT *Il1b* haplotype presenting markedly lower levels of *Tgfb1* expression.

5.4.10 Inflammatory and anti-inflammatory responses

In order to examine the possibility that a genetic variant could affect the immune response as a whole, single response variables describing either a general pro- or anti-inflammatory response were derived using PCA, to group sets of positively correlated, functionally related variables (Table 5.4) (Jackson *et al.* in review). The inflammatory variable PC^{I} had an eigenvalue of 2.86 and explained 35.8% of the total variation, while the anti-inflammatory variable PC^{AI} explained 26.3% of the total variation and returned an eigenvalue of 2.10. The minimal non-genetic model for the grouped inflammatory variable was: PC^{I} ~ season (Table 5.2). Results of the model indicated that PC^{I} expression was low in summer and autumn and elevated in winter and spring 2009 (Appendix Table 2.5). For the anti-inflammatory variable the minimal model was defined as: PC^{AI} ~ season + body weight + eye lens weight + sex + season × sex + body weight × eye lens weight. Elevated expression of PC^{AI} was observed in winter, although males in winter were more likely to exhibit lower anti-inflammatory responses (Appendix Table 2.5). No genetic terms were associated with either of the two more general immune response variables.

Table 5.4 Component loadings of individual variables for first principal component of pro- and anti-inflammatory expression variables

PC^{I}; inflammatory		PC^{AI}; anti-inflammatory	
Component	Coefficient	Component	Coefficient
Ifng 96 h	0.432	*Tgfb1* 24 h TLR2	0.219
Ifng 96 h PHA	0.433	*Il10* 24h TLR7	0.369
Il1b 24 h	0.223	*Il10* 24h	0.393
Il1b 24 h TLR2	0.125	*Tgfb1* 24 h	0.374
Il2 96 h	0.361	*Foxp3* 96 h PHA	0.444
Il2 96h PHA	0.331	*Il10* 96 h PHA	0.156
Tbx21 96 h	0.374	*Tgfb1* 96 h PHA	0.523
Tbx21 96 h PHA	0.426	*Tgfb1* 96 h	0.160

5.5 DISCUSSION

Functional genetic variation can have an effect on an individual's phenotype through either structural changes to the translated protein or through altered transcription levels. In previous chapters, the field vole genes *Il1b* and *Il2* have exhibited strong evidence that genetic diversity has been maintained at both loci through parasite-driven selection, while there is also evidence that polymorphism within *Il12b* is functional. In the present study, polymorphism within all three genes was associated with variation in host immune phenotype, as measured through expression of immunity-related genes. Specific associations with several genes were observed, including, in the case of *Il1b*, a strong association with its own expression level. Furthermore, polymorphism at the *Il12b* locus was strongly associated with expression of both *Il1b* and *Il2*, suggesting that the associations with parasite resistance observed for these genes may be explained by altered expression levels mediated through the cytokine network.

5.5.1 Genetic associations with immune phenotype

This chapter examined the hypothesis that variation in immune function is the phenotypic mechanism underlying the consistent associations between the genes *Il1b, Il2* and *Il12b* with parasite resistance, by examining associations between genetic polymorphism and expression levels of several immune genes. Although it

should be stated that measuring the levels of immune gene expression in artificially-stimulated cells does not necessarily correlate

Polymorphism at the *Il1b* locus was strongly associated with its own mRNA expression. This was the only gene where variation within the locus was associated with differential expression of that same gene. It is unlikely that the polymorphisms typed within *Il1b* in this study could themselves affect transcription, as they are located in the exonic regions of the gene. It is more likely that the causative mutation is an un-typed *cis*-acting polymorphism in linkage disequilibrium to the genotyped SNPs, particularly as LD is strong within this gene (Chapter 3). In humans, *IL1B* mRNA expression has been shown to be a heritable trait and polymorphisms within the promoter region (in particular the -31 T>C SNP) are thought to be functional (Smith and Humphries 2008). It is possible that the field vole *Il1b* gene contains a polymorphism of similar effect; identifying the precise causative mutation and the molecular mechanism of altered gene expression is however beyond the scope of this study. The *Il1b* locus was also associated with 96 hour unstimulated *Gata3* expression, and near-significant associations were observed between *Il1b* and mRNA accumulations of *Ifng* and *Tgfb1*, providing evidence that variation within *Il1b* can have discernible effects on the host's immune phenotype.

IL-1β is an important, pleiotropic mediator of the innate response and a potent pro-inflammatory cytokine. The multi-functional nature of IL-1β means that is perhaps not surprising that genetic polymorphism within may have a discernible effect on several other genes; if variant alleles lead to differential expression then downstream events of the immune response may be affected (Kelso 1998). As polymorphism within *Il1b* appears to have an effect on its own expression levels, it is possible that the associations with other genes are related to this altered transcription. *Il1b* expression was, when fitted as an additional explanatory variable, very strongly associated with *Tgfb1* expression levels ($p < 0.0001$), for example. However, the complexity of the cytokine network and the vertebrate immune system as a whole means that the precise pathways and mechanisms by

which *Il1b* polymorphism appears to influence expression of other genes in this natural field vole population are, as yet, unclear.

Il2 polymorphism was significantly associated with expression of *Il10* (and marginally non-significant with *Il1b*) but not with its own transcription levels. The interaction between *Il2* variation and expression of *Il10* is then perhaps likely to be somewhat more subtle than a correlation in expression levels between the two genes. The two cytokines are known to interact at the protein level; IL-10 is an anti-inflammatory cytokine which inhibits the production of many other cytokines involved in Th1, Th2 and pro-inflammatory responses, including IL-2 (Borish and Steinke 2003; Steinke and Borish 2006). There is then scope for a genetic basis influencing this relationship, but further work would be needed to elucidate the precise mechanism of such an interaction.

There were strong associations between the *Il12b* locus and expression of both *Il1b* and *Il2*. Importantly, these three genes were those that were most consistently associated with parasite resistance in the previous chapter and those for which the evidence of functional variation is most compelling. *Il1b* and *Il2* have previously exhibited strong evidence for the action of balancing selection maintaining genetic diversity at both loci (Chapters 2 and 3), and for associations with pathogen resistance (Chapter 4); *Il12b* has also been associated with parasite susceptibility (Chapter 3) and contains a SNP predicted to have an effect on the function of the translated protein and which is segregating at very low frequencies (Chapters 2 and 3). That these same genes are the most consistently associated with gene expression accumulates more support for the functional relevance of the genetic variation within these loci. IL-1β, IL-2 and IL-12 are all broadly immunostimulatory cytokines which through acting on the same cells interact with one another during both the innate and adaptive immune response, with IL-12 acting as a bridge between these two arms of the immune system (Romani *et al.* 1997; Borish and Steinke 2003). It is interesting that in this study polymorphism within each of these three genes was associated with expression of at least one of the others (Figure 5.1), which may suggest a possible shared physiological pathway in which interactions

may occur *in vivo*. The IL-12 p40 subunit encoded by *Il12b* also forms part of the related cytokine IL-23, which is another mediator of the inflammatory response and is involved in the production of Th17 cells (Weaver *et al.* 2006; Kastelein *et al.* 2007). There are then a number of possible pathways in which *Il1b*, *Il2* and *Il12b* may interact *in vivo* and the precise mechanisms or directionality underlying any such interactions within the field vole are unknown. However, the consistency of associations between these three genes across several complementary research approaches in this and previous chapters provides support for the presence of causal relationships, and that genetic variation within immune genes can give rise to phenotypic effects mediated through the cytokine network.

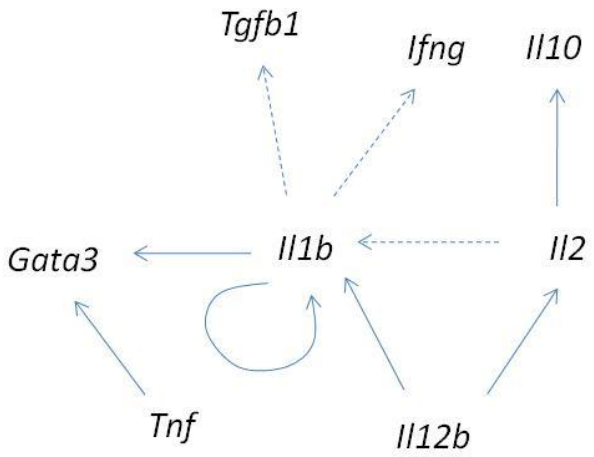

Figure 5.1 Summary diagram of genetic associations with gene expression. Arrows indicate where polymorphism within one gene is associated with expression levels in another. No directionality of *in vivo* molecular pathways is implied. Solid arrows represent statistically significant associations ($p < 0.05$), dashed arrows represent marginally non-significant associations ($0.05 < p < 0.07$).

The lack of genetic associations with the more general inflammatory and anti-inflammatory variables probably reflects the complexity of the vertebrate immune response. Whereas polymorphism in one gene may affect single genes involved in similar biochemical pathways, such as the example of *Il1b* and *Il2*, there may be less power to detect associations with a wider, more composite measure of immune responsiveness. The polygenic, dynamic and tightly regulated nature of the immune system may mean it is unlikely that one will find a single genetic polymorphism able to affect the immune system in such a wide-spread manner.

Previous studies have indicated that much of the genetic component underlying pathogen resistance is due to many minor susceptibility genes, rather than a few major loci (Hill 2001b). It is then more likely that, in the field vole too, a wide range of polymorphisms in many genes, each with relatively subtle effects, will contribute to the immune response each time it is invoked.

5.6 CONCLUSION

Identifying functional genetic diversity in immune-related genes is of clear importance. However, studies designed to understand the mechanisms by which polymorphism in immune genes of natural populations can give rise to phenotypic effects have until now been largely neglected. This study provides evidence that polymorphism within immune genes of a wild field vole population can lead to both *cis*- and *trans*-acting effects on gene expression. Variation in host immune function therefore provides a possible explanatory mechanism for the genetic basis of parasite resistance in the field vole.

Chapter 6

General discussion

6.1 SUMMARY

The broad aim of this thesis was to expand research on wildlife immunogenetics away from genes of the MHC, in order to increase our understanding of the role of natural selection in shaping the genetic diversity of the immune system. This study examined polymorphism within a wide range of cytokine and other non-MHC genes in a natural population. I have accumulated strong evidence that natural selection has driven the maintenance of genetic diversity within a subset of field vole cytokine genes, particularly *Il1b* and *Il2*, for which polymorphisms were also consistently associated with variation in immune phenotype and in pathogen susceptibility. Hence, for the first time, I have demonstrated that parasites can act as a selective force to maintain diversity at non-MHC immune genes in natural populations, and that variation in immune function, mediated through the cytokine network, is likely to contribute to the genetic basis of parasite resistance in wild field voles.

The central role of cytokines in immunity has led to tens of thousands of published studies into their structure and function, of which the overwhelming majority have been performed on humans or laboratory rodents. Laboratory studies have provided valuable mechanistic insights into the role of cytokines in regulating the immune response, while human studies have highlighted the functional consequences of cytokine polymorphism - i.e. associations with infectious and non-communicable diseases - and the evolutionary pressures that may have shaped cytokine genetic diversity (Daser *et al.* 1996; Kelso 1998; Van Deventer 2000; O'Shea *et al.* 2002; Borish and Steinke 2003; Ollier 2004; Hollegaard and Bidwell 2006; Steinke and Borish 2006; Fumagalli *et al.* 2009; Smith and Humphries 2009).

However, despite the crucial role of cytokines in the immune response, there has been very little research into how natural selection shapes cytokine diversity in natural populations, or the phenotypic consequences of this variation. A major advantage of studying natural populations is that it allows us to detect evolution in action; this has been demonstrated ably by the many studies into the evolution of the MHC in wild vertebrates (reviewed in Piertney and Oliver 2006). However, genes of the MHC account for only fraction of the variation observed in disease susceptibility (Jepson *et al.* 1997) and so by broadening immunogenetic research to include cytokines and other non-MHC genes in wild species, we can improve our understanding of the levels at which natural selection can act on the host immune system (Acevedo-Whitehouse and Cunningham 2006). The work of this thesis therefore aimed to bridge the current gap in our knowledge between the mechanistic understanding of cytokine biology garnered through studies of laboratory rodents and the variation in immune function and disease susceptibility observed in nature.

In Chapters 2 and 3 I have shown that the patterns of genetic diversity observed in several immune genes within a natural population of field voles are likely to have been shaped by natural selection. The strongest signatures of selection were observed within the genes *Il1b*, *Il2* and *Tnf*, where patterns of polymorphism were consistent with the action of balancing selection maintaining genetic diversity. Importantly, these inferences were consistent over several complementary approaches, including tests of neutrality at both individual DNA sequence and population levels, and examination of allele frequencies within and between several subpopulations. A combination of approaches such as this is important as it allows the detection of signatures of selection over several different timescales. Comparison of DNA sequence diversity between species can provide insights into selection occurring over evolutionary timescales and the pressures that have driven species divergence. Conversely, examination of allele frequencies within and between populations enables us to detect signatures of selection occurring over relatively recent timescales (Beaumont 2005; Excoffier *et al.* 2009). At the DNA sequence level, *Il1b*, *Il2* and *Tnf* all exhibited high levels of genetic diversity and an excess of intermediate frequency mutations (Chapter 2) while, at the population

level, these loci displayed an excess of heterozygosity and more even haplotype frequencies than one would expect under neutrality (Chapter 3). SNPs within *Il1b* and *Il2* also demonstrated increased levels of differentiation between populations compared to other loci (Chapter 3). Taken together, these findings provide strong corroborative evidence that balancing selection has been an important force in maintaining the high levels of genetic diversity observed within these field vole cytokine genes. Balancing selection at immune loci is well documented in humans (Ferrer-Admetlla *et al.* 2008; Fumagalli *et al.* 2009; Barreiro and Quintana-Murci 2010) and in wildlife, particularly at the MHC (Piertney and Oliver 2006; Jensen *et al.* 2008; Tennessen and Blouin 2008), but this is the first study to demonstrate that selection can act to maintain high levels of diversity within cytokine genes of a free-living natural population.

The analyses of chapters 2 and 3 also revealed some evidence for positive directional selection driving the evolution of the genes encoding the *Ifng* cytokine and the pattern-recognition receptor, *Tlr4*. Genetic diversity within the field vole *Ifng* gene was low, as has been observed in several other species (Worley *et al.* 2006; Downing *et al.* 2009). Between-species divergence was, however, the highest of all genes examined (Chapter 2). These patterns may therefore be indicative of past episodes of positive selection driving the genetic divergence of *Ifng* between species, while reducing the level of diversity around the site of selection within the field vole, which is a characteristic signature of selective sweeps and genetic hitchhiking (Maynard Smith and Haigh 1974; Nielsen 2005). This suggestion is supported by the recent findings of Levi-Acobas *et al*. (2009), who found evidence for several episodes of strong, positive Darwinian selection acting on the *Ifng* gene in the Glires lineage, a mammalian clade of which voles are a member. However, their study relied on between-species comparisons of DNA sequences derived from laboratory or wild-derived strains and did not therefore utilise data from natural populations.

The field vole *Tlr4* gene also exhibited evidence of positive selection. Although the results of sequence-based neutrality tests were borderline non-significant ($0.05 < p$

< 0.06), test statistics were consistently negative across several types of test (Chapter 2), which can indicate the action of directional selection. Furthermore, a signature of positive selection was identified via the Ewens-Watterson test (Ewens 1972; Watterson 1978), which demonstrated that the most common *Tlr4* haplotype is segregating at higher frequencies than one would expect under a model of neutral evolution (Chapter 3). An F_{ST}-based outlier test (Excoffier *et al.* 2009) identified a nonsynonymous SNP in the *Tlr4* locus (Tlr4 1663 A/G) as a candidate for positive selection. At this site, the derived *A* allele is segregating at high frequencies in all subpopulations and has almost reached fixation within Kielder Forest (99% frequency), which may be indicative of selection acting to drive the increase of the comparatively recent *A* allele.

In order to verify inferences of selection acting on specific genes it is important to demonstrate that genetic diversity has phenotypic consequences, as natural selection acts on phenotypic rather than genotypic variation. In Chapter 4 I examined the functional and phenotypic significance of the immunogenetic polymorphism observed previously, under the hypothesis that parasites are the selective force that` has shaped the genetic diversity of field vole immune genes. Genetic variation in two genes in particular, *Il1b* and *Il2*, was consistently associated with variation in resistance to a diverse range of pathogens. As these two genes exhibited strong evidence for balancing selection in the previous chapters, linking the evidence for selection and associations with infectious disease susceptibility provides robust support for the idea that parasite-driven selection underpins the maintenance of genetic variation within *Il1b* and *Il2*. Several mechanisms have been proposed as to how pathogen-driven selection can drive the maintenance of immunogenetic diversity, including overdominance (Doherty and Zinkernagel 1975), negative frequency-dependent selection (Bodmer 1972) and fluctuating selection (Hedrick 2002). As allele frequencies at both *Il1b* and *Il2* varied between neighbouring subpopulations, and abundances of all parasites varied over time or space (or both), fluctuating selection and local adaptation is perhaps the most likely explanatory mechanism for the balanced diversity observed at the *Il1b* and *Il2* loci. The lack of consistent associations between the other strong candidate gene for balancing

selection, *Tnf*, and pathogen susceptibility may indicate that parasites are not the main selective force shaping diversity at this locus, or that parasite species unaccounted for in this study are more important. Parasite-driven selection has been a major force in the evolution of immune genes, including the MHC in wildlife (Paterson *et al.* 1998; Sommer 2005; Piertney and Oliver 2006) and interleukins in humans (Fumagalli *et al.* 2009); the work of this thesis has uniquely demonstrated that molecular evolution of cytokines in natural populations may also be driven through selection acting on phenotypic variation in infectious disease resistance.

In Chapter 5 I demonstrated that polymorphisms within *Il1b* and *Il2* are also associated with variation in immune function, as measured through expression of several different cytokines and transcription factors. This observation provides further evidence of the phenotypic consequences of polymorphism within these genes and suggests that variation in host immune phenotype may be the mechanism underlying the genetic association with parasite susceptibility observed at these loci. Expression levels of both these genes were influenced by an individual's *Il12b* genotype. Polymorphism within *Il12b* was also associated with resistance to a number of parasites (Chapter 4) and was previously predicted to affect protein function (Chapter 2). Although without further experimental work one cannot identify the interactions that may occur between these three genes *in vivo*, all are involved in the mediation of the pro-inflammatory response and may therefore share similar physiological pathways. The accumulated evidence for phenotypic consequences of polymorphism at these loci provides strong support for the presence of true causal relationships, and that genetic variation within immune genes of wild field voles can give rise to phenotypic effects mediated through the cytokine network.

By testing multiple response variables and examining signatures of selection over several different biological levels (individual, population, species) and timescales, one avoids highlighting single genetic associations which may arise from type I errors and increases the reliability of any resulting biological inferences.

Throughout this thesis I have therefore particularly emphasised those genes that have demonstrated signatures of selection across several types of test and/or exhibited repeated associations with pathogen resistance and immune function in independent datasets, of which *Il1b* and *Il2* were the most consistent.

6.2 WILD RODENTS AS MODELS IN EVOLUTIONARY ECOLOGY

More broadly, this thesis highlights the benefits of utilising wild rodents as a model system for studying natural selection, parasite resistance and immune function. Traditionally, studies examining how and why individuals vary in immune function and disease susceptibility have concentrated on human or laboratory model species, both of which are associated with difficulties. Large-scale studies of human populations are logistically challenging, labour-intensive and time-consuming, particularly in the developing world. It is also not possible to sample invasively from human populations, and as such it is impossible to obtain tissues, such as spleen or gut, which one may require for immunological assays or accurate parasite diagnoses. As an alternative, laboratory rodents have been used extensively in studies of immunology, and have provided a tremendous amount of information on the cellular and molecular basis of the immune response. However, laboratory species are inbred and are not representative of natural genetic variation. Furthermore, such animals are domesticated, kept in stable conditions under little or no climatic or nutritional stresses and without the exposure to multiple pathogens characteristic of natural populations. Such studies therefore do not offer any insight into the role of natural selection or environmental heterogeneity on the immune response and are therefore divorced from an ecological context (Grenfell *et al.* 2002).

By studying wild rodent populations we have the opportunity to extend laboratory immunology to the field and to examine the immune response within the context in which it has evolved. Through the identification of genetic factors which contribute to variation in immune function and disease susceptibility, this thesis has highlighted the potential of using cytokines to measure immune function in natural populations, beyond the simplified view of 'immunocompetence' (Graham *et al.*

2007). By taking into account the various arms of the immune response we can better define the immune status of hosts (Bradley and Jackson 2008) and hence increase our understanding of how host immune phenotype influences life-history traits, co-infection, and susceptibility to parasites (Graham *et al.* 2007; Bradley and Jackson 2008; Jackson *et al.* in review). Wild rodents have great potential as a new model for studying immunology in an ecologically valid context, as they are amenable to large-scale controlled experiments in the field that would not be possible in humans or endangered species. In addition, wild-derived individuals could be bred in captivity to more closely examine the effect of different parasite loads, genotypes and environmental conditions on host immunology and fitness, using experiments analogous to those conducted on laboratory mice.

The advent of next-generation sequencing means it is now possible to identify hundreds or thousands of SNPs simultaneously (Metzker 2010), including in wild, non-model organisms (Ellegren 2008; Vera *et al.* 2008; Slate *et al.* 2009; van Bers *et al.* 2010). There are several high-throughput platforms capable of genotyping at this scale (Ding and Jin 2009) and, although in the main these have been developed for model organisms or humans, the technology has the potential to be used on any organism (Slate *et al.* 2009). It is at present still difficult, and costly, to type large numbers of SNPs in a non-model organism. Molecular ecologists have therefore had to decide whether to sequence a large number of SNPs in a few individuals, or to genotype a large number of individuals at relatively few SNPs. As technology develops and costs decrease it will however become increasingly feasible for large-scale genomic research to be undertaken on natural populations, and thus allow genomic data to be analysed in an ecological and evolutionary context (Tautz *et al.* 2010). It will soon become possible for the first time to carry out genome-wide association studies on any natural populations with good quality phenotypic data (Slate *et al.* 2010). Traditionally, gene mapping studies in the wild have been conducted through typing of microsatellite markers identified in closely related, often domesticated, organisms. Using SNPs identified through high-throughput sequencing methods as genetic markers offers several advantages over the use of microsatellites: development of SNP markers is less time consuming, typing can be

automated and performed on a much larger scale and, as SNPs are much more abundant, any genome location can be analysed (Slate *et al.* 2009).
The next logical step for immunogenetic research on the field vole is therefore to expand on the work presented in this thesis and utilise genomic technology to study hundreds, rather than dozens, of immunity genes. By sequencing and typing hundreds or thousands of SNPs in a large number of individuals we could potentially examine the genetic diversity of the full range of field vole immune genes. This would provide a genome scan to identify which genes are under selection, and thus allow for a much more comprehensive evaluation of the role of natural selection on immunity. It would no doubt prove fruitful to generate genomic data for several field vole populations and for other wild rodent species too; this would allow one to not only identify which immune genes are under selection in the field vole, but to ascertain whether those same genes are under selection in different populations or species. One advantage of using wild rodents as a model for this kind of study is that the house mouse, *Mus musculus*, has a complete genome sequence and extensive information on immune gene function; this would greatly facilitate development of SNP-typing methods and the identification of orthologous genes in related species, as it has in the present study. By comparing DNA sequence diversity between closely related species (for example field and bank voles, or wood and house mice) we may also gain new insights into the extent to which balancing and directional selection have contributed to immunogenetic diversity within species and driven species divergence (Hudson *et al.* 1987; McDonald and Kreitman 1991; Obbard *et al.* 2009).

In chapters 3 and 4, I demonstrated the use of GLMs and GLMMs in controlling for non-genetic factors and analysing the extent to which genes under selection exhibit phenotypic variation in parasite resistance and immune function; such analyses would be easily transferable to the more comprehensive studies proposed here. Variation in parasite resistance was, in this thesis, defined as the probability of acquiring infection; a genotype may be seen to confer resistance by decreasing this probability. However, a genotype may also confer resistance by shortening the

duration of infection or increasing the chance of surviving an infection. These possibilities could be explored through the use of longer-term capture-recapture studies, involving life-long tracking of tagged and genotyped individuals. Such studies would broaden our understanding of the genetic basis of parasite susceptibility by allowing us to ask questions such as, for a given genotype, when does an animal become infected? How long does it take to clear the infection? Is the protective genotype detrimental in the absence of infection? And, what are the consequences of infection for host survival and fitness?

In summary, field voles and other wild rodents are an extremely promising system for studying selection in natural populations, and for bridging the gap between the mechanistic knowledge garnered from laboratory mice and the variation in immune function and disease resistance observed in humans and wildlife. This thesis demonstrates the potential of such studies, especially given new sequencing technologies that will increase the ease with which genomic resources can be transferred from laboratory to natural populations.

References

Abbas, A. K., Murphy, K. M. and Sher, A. (1996). "Functional diversity of helper T lymphocytes." Nature **383**(6603): 787-793.

Acevedo-Whitehouse, K. and Cunningham, A. A. (2006). "Is MHC enough for understanding wildlife immunogenetics?" Trends in Ecology and Evolution **21**(8): 433-438.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. International Symposium on Information Theory. B. N. Petran and F. Csaki. Budapest, Akademiai Kiadi**:** 267-281.

Akashi, H. (1995). "Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in Drosophila DNA." Genetics **139**(2): 1067-1076.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). "Basic local alignment search tool." Journal of Molecular Biology **215**(3): 403-410.

Balkwill, F. (2002). Cytokines and cytokine receptors. Immunology. I. Roitt, J. Brostoff and D. Male. London, Mosby**:** 119-129.

Barreiro, L. B. and Quintana-Murci, L. (2010). "From evolutionary genetics to human immunology: How selection shapes host defence genes." Nature Reviews Genetics **11**(1): 17-30.

Bayley, J. P., Ottenhoff, T. H. M. and Verweij, C. L. (2004). "Is there a future for TNF promoter polymorphisms?" Genes and Immunity **5**(5): 315-329.

Beaumont, M. A. (2005). "Adaptation and speciation: what can Fst tell us?" Trends in Ecology & Evolution **20**(8): 435-440.

Beaumont, M. A. and Balding, D. J. (2004). "Identifying adaptive genetic divergence among populations from genome scans." Molecular Ecology **13**(4): 969-980.

Behnke, J. M., Iraqi, F., Menge, D., Baker, R. L., Gibson, J. and Wakelin, D. (2003). "Chasing the genes that control resistance to gastrointestinal nematodes." Journal of Helminthology **77**(2): 99-109.

Bell, G. (2010). "Fluctuating selection: The perpetual renewal of adaptation in variable environments." Philosophical Transactions of the Royal Society B: Biological Sciences **365**(1537): 87-97.

Benton, M. J. and Donoghue, P. C. J. (2007). "Paleontological evidence to date the tree of life." Molecular Biology and Evolution **24**(1): 26-53.

Birtles, R. J., Hazel, S. M., Bennett, M., Bown, K., Raoult, D. and Begon, M. (2001). "Longitudinal monitoring of the dynamics of infections due to Bartonella species in UK woodland rodents." Epidemiology and Infection **126**(2): 323-329.

Biswas, S. and Akey, J. M. (2006). "Genomic insights into positive selection." Trends in Genetics **22**(8): 437-446.

Black, W. C., Baer, C. F., Antolin, M. F. and DuTeau, N. M. (2001). Population genomics: Genome-wide sampling of insect populations. Annual Review of Entomology. **46:** 441-469.

Black, W. C. and Krafsur, E. S. (1985). "A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients." TAG Theoretical and Applied Genetics **70**(5): 491-496.

Blackwell, J. M., Goswami, T., Evans, C. A. W., Sibthorpe, D., Papo, N., White, J. K., Searle, S., Miller, E. N., Peacock, C. S., Mohammed, H. and Ibrahim, M. (2001). "SLC11A1 (formerly NRAMP1) and disease resistance." Cellular Microbiology **3**(12): 773-784.

Bluestone, J. A. and Abbas, A. K. (2003). "Natural versus adaptive regulatory T cells." Nature Reviews Immunology **3**(3): 253-257.

Bodmer, W. F. (1972). "Evolutionary Significance of the HL-A System." Nature **237**(5351): 139-183.

Borish, L. C. and Steinke, J. W. (2003). "2. Cytokines and chemokines." Journal of Allergy and Clinical Immunology **111**(2 SUPPL. 2).

Bown, K. J., Begon, M., Bennett, M., Birtles, R. J., Burthe, S., Lambin, X., Telfer, S., Woldehiwet, Z. and Ogden, N. H. (2006). "Sympatric ixodes trianguliceps and ixodes ricinus ticks feeding on field voles (Microtus agrestis): Potential for increased risk of Anaplasma phagocytophilum in the United Kingdom?" Vector-Borne and Zoonotic Diseases **6**(4): 404-410.

Bown, K. J., Begon, M., Bennett, M., Woldehiwet, Z. and Ogden, N. H. (2003). "Seasonal dynamics of *Anaplasma phagocytophila* in a rodent-tick (*Ixodes trianguliceps*) system, United Kingdom." Emerging Infectious Diseases **9**(1): 63-70.

Bown, K. J., Lambin, X., Telford, G. R., Ogden, N. H., Telfer, S., Woldehiwet, Z. and Birtles, R. J. (2008). "Relative importance of *Ixodes ricinus* and *Ixodes trianguliceps* as vectors of *Anaplasma phagocytophilum* and *Babesia microti* in field vole (*Microtus agrestis*) populations." Applied and Environmental Microbiology **74**(23): 7118-7125.

Box, G. E. P. and Cox, D. R. (1964). "An analysis of transformations (with discussion)." Journal of the Royal Statistical Society B **26**: 211–252.

Bradley, J. E. and Jackson, J. A. (2008). "Measuring immune system variation to help understand host-pathogen community dynamics." Parasitology **135**(Special Issue 07): 807-823.

Burthe, S., Bennett, M., Kipar, A., Lambin, X., Smith, A., Telfer, S. and Begon, M. (2008a). "Tuberculosis (*Mycobacterium microti*) in wild field vole populations." Parasitology **135**(3): 309-317.

Burthe, S., Telfer, S., Begon, M., Bennett, M., Smith, A. and Lambin, X. (2008b). "Cowpox virus infection in natural field vole *Microtus agrestis* populations: significant negative impacts on survival." Journal of Animal Ecology **77**(1): 110-119.

Burthe, S., Telfer, S., Lambin, X., Bennett, M., Carslake, D., Smith, A. and Begon, M. (2006). "Cowpox virus infection in natural field vole *Microtus agrestis populations*: Delayed density dependence and individual risk." Journal of Animal Ecology **75**(6): 1416-1425.

Cavalli-Sforza, L. L. (1966). "Population structure and human evolution." Proceedings of the Royal Society B: Biological Sciences **164**(995): 362-379.

Cavanagh, R., Begon, M., Bennett, M., Ergon, T., Graham, I. M., De Haas, P. E. W., Hart, C. A., Koedam, M., Kremer, K., Lambin, X., Roholl, P. and Van

Soolingen, D. (2002). "*Mycobacterium microti* infection (vole tuberculosis) in wild rodent populations." Journal of Clinical Microbiology **40**(9): 3281-3285.

Cavanagh, R. D., Lambin, X., Ergon, T., Bennett, M., Graham, I. M., Van Soolingen, D. and Begon, M. (2004). "Disease dynamics in cyclic populations of field voles (*Microtus agrestis*): Cowpox virus and vole tuberculosis (*Mycobacterium microti*)." Proceedings of the Royal Society B: Biological Sciences **271**(1541): 859-867.

Charbonnel, N., Goüy de Bellocq, J. and Morand, S. (2006). Immunogenetics of micromammalmacroparasite interactions. Micromammals and Macroparasites**:** 401-442.

Charlesworth, D. (2006). "Balancing Selection and Its Effects on Sequences in Nearby Genome Regions." PLoS Genetics **2**(4): e64.

Clark, A. G. (2004). "The role of haplotypes in candidate gene studies." Genetic Epidemiology **27**(4): 321-333.

Cleaveland, S., Hess, G. R., Dobson, A. P., Laurenson, M. K., McCallum, H. I. and Roberts, M. G. (2002). The role of pathogens in biological conservation. The Ecology of Wildlife Diseases. P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heesterbeek and A. P. Dobson. Oxford, Oxford University Press**:** 139–150.

Clutterbuck, E. J., Hirst, E. M. A. and Sanderson, C. J. (1989). "Human interleukin-5 (IL-5) regulates the production of eosinophils in human bone marrow cultures: Comparison and interaction with IL-1, IL-3, IL-6, and GMCSF." Blood **73**(6): 1504-1512.

Coltman, D. W., Wilson, K., Pilkington, J. G., Stear, M. J. and Pemberton, J. M. (2001). "A microsatellite polymorphism in the gamma interferon gene is associated with resistance to gastrointestinal nematodes in a naturally-parasitized population of Soay sheep." Parasitology **122**(5): 571-582.

Cooper, M. D. and Alder, M. N. (2006). "The Evolution of Adaptive Immune Systems." Cell **124**(4): 815-822.

Corbet, G. and Harris, S., Eds. (1991). The Handbook of British Mammals. Oxford, Blackwell.

Corwin, E. J. (2000). "Understanding cytokines. Part I: Physiology and mechanism of action." Biological research for nursing **2**(1): 30-40.

Cote-Sierra, J., Foucras, G., Guo, L., Chiodetti, L., Young, H. A., Hu-Li, J., Zhu, J. and Paul, W. E. (2004). "Interleukin 2 plays a central role in Th2 differentiation." Proceedings of the National Academy of Sciences of the United States of America **101**(11): 3880-3885.

Dalziel, A. C., Rogers, S. M. and Schulte, P. M. (2009). "Linking genotypes to phenotypes and fitness: How mechanistic biology can inform molecular ecology." Molecular Ecology **18**(24): 4997-5017.

Daser, A., Mitchison, H., Mitchison, A. and Müller, B. (1996). "Non-classical-MHC genetics of immunological disease in man and mouse. The key role of pro-inflammatory cytokine genes." Cytokine **8**(8): 593-597.

Dean, M., Carrington, M. and O'Brien, S. J. (2002). "Balanced polymorphism selected by genetic versus infectious human disease." Annual Review of Genomics and Human Genetics **3**: 263-292.

Devevey, G. and Christe, P. (2009). "Flea infestation reduces the life span of the common vole." Parasitology **136**(11): 1351-1355.

Devevey, G., Niculita-Hirzel, H., Biollaz, F., Yvon, C., Chapuisat, M. and Christe, P. (2008). "Developmental, metabolic and immunological costs of flea infestation in the common vole." Functional Ecology **22**(6): 1091-1098.

Ding, C. and Jin, S. (2009). "High-throughput methods for SNP genotyping." Methods in Molecular Biology (Clifton, N.J.) **578**: 245-254.

Doherty, P. C. and Zinkernagel, R. M. (1975). "Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex." Nature **256**(5512): 50-52.

Downing, T., Lloyd, A. T., O'Farrelly, C. and Bradley, D. G. (2010). "The Differential Evolutionary Dynamics of Avian Cytokine and TLR Gene Classes." Journal of Immunology **184**(12): 6993-7000.

Downing, T., Lynn, D., Connell, S., Lloyd, A., Bhuiyan, A. K., Silva, P., Naqvi, A., Sanfo, R., Sow, R.-S., Podisi, B., O'Farrelly, C., Hanotte, O. and Bradley, D. (2009). "Contrasting evolution of diversity at two disease-associated chicken genes." Immunogenetics **61**(4): 303-314.

Ellegren, H. (2008). "Sequencing goes 454 and takes large-scale genomics into the wild." Molecular Ecology **17**(7): 1629-1631.

Else, K. and Wakelin, D. (1988). "The effects of H-2 and non-H-2 genes on the expulsion of the nematode Trichuris muris from inbred and congenic mice." Parasitology **96**.

Else, K. and Wakelin, D. (1989). "Genetic variation in the humoral immune responses of mice to the nematode Trichuris muris." Parasite Immunology **11**(1): 77-90.

Else, K. J., Finkelman, F. D., Maliszewski, C. R. and Grencis, R. K. (1994). "Cytokine-mediated regulation of chronic intestinal helminth infection." Journal of Experimental Medicine **179**(1): 347-351.

Elton, C., Davis, D. H. S. and Findlay, G. M. (1935). "An Epidemic among Voles (Microtus agrestis) on the Scottish Border in the Spring of 1934." Journal of Animal Ecology **4**(2): 277-288.

Ewens, W. J. (1972). "The sampling theory of selectively neutral alleles." Theoretical Population Biology **3**(1): 87-112.

Excoffier, L., Hofer, T. and Foll, M. (2009). "Detecting loci under selection in a hierarchically structured population." Heredity.

Excoffier, L., Laval, G. and Schneider, S. (2005). "Arlequin (version 3.0): An integrated software package for population genetics data analysis." Evolutionary Bioinformatics Online **1**: 47-50.

Excoffier, L. and Lischer, H. E. L. (2010). "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." Molecular Ecology Resources **10**(3): 564-567.

Farrar, M. A. and Schreiber, R. D. (1993). "The molecular cell biology of interferon-γ and its receptor." Annual Review of Immunology **11**: 571-611.

Fay, J. C. and Wu, C. I. (2000). "Hitchhiking under positive Darwinian selection." Genetics **155**(3): 1405-1413.

Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marquès-Bonett, T., Ramírez-Soriano, A., Muntasell, A., Navarro, A., Lazarus, R., Calafell, F., Bertranpetit, J. and Casals, F. (2008). "Balancing selection is the main force shaping the

evolution of innate immunity genes." Journal of Immunology **181**(2): 1315-1322.

Flajnik, M. F. and Du Pasquier, L. (2004). "Evolution of innate and adaptive immunity: can we draw a line?" Trends in Immunology **25**(12): 640-644.

Freeman, A., Lynn, D., Murray, C. and Bradley, D. (2008). "Detecting the effects of selection at the population level in six bovine immune genes." BMC Genetics **9**(1): 62.

Fu, Y. X. and Li, W. H. (1993). "Statistical Tests of Neutrality of Mutations." Genetics **133**(3): 693-709.

Fumagalli, M., Pozzoli, U., Cagliani, R., Comi, G. P., Riva, S., Clerici, M., Bresolin, N. and Sironi, M. (2009). "Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions." Journal of Experimental Medicine **206**(6): 1395-1408.

Garnier-Gere, P. and Dillmann, C. (1992). "A computer program for testing pairwise linkage disequilibria in subdivided populations." Journal of Heredity **83**(3): 239.

Gomez, L. M., Camargo, J. F., Castiblanco, J., Ruiz-Narváez, E. A., Cadena, J. and Anaya, J. M. (2006). "Analysis of IL1B, TAP1, TAP2 and IKBL polymorphisms on susceptibility to tuberculosis." Tissue Antigens **67**(4): 290-296.

Graham, A. L., Allen, J. E. and Read, A. F. (2005). Evolutionary causes and consequences of immunopathology. Annual Review of Ecology, Evolution, and Systematics. **36:** 373-397.

Graham, A. L., Cattadori, I. M., Lloyd-Smith, J. O., Ferrari, M. J. and Bjørnstad, O. N. (2007). "Transmission consequences of coinfection: cytokines writ large?" Trends in Parasitology **23**(6): 284-291.

Grencis, R. K. (2001). "Cytokine regulation of resistance and susceptibility to intestinal nematode infection - From host to parasite." Veterinary Parasitology **100**(1-2): 45-50.

Grenfell, B. T., Amos, W., Arneberg, P., Bjørnstad, O. N., Greenman, J. V., Harwood, J., Lanfranchi, P., McLean, A. R., Norman, R. A., A.F.S, R. and Skorping, A. (2002). Visions for future research in wildlife epidemiology. The Ecology of Wildlife Diseases. P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heesterbeek and A. P. Dobson. Oxford, Oxford University Press**:** 151-164.

Guo, S. W. and Thompson, E. A. (1992). "Performing the exact test of Hardy-Weinberg proportion for multiple alleles." Biometrics **48**(2): 361-372.

Haldane, J. B. S. (1949). "Disease and Evolution." La Ricerca Scientifica. Supplement A. **19**: 68–76.

Hall, T. A. (1999). "BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." Nucleic Acids Symposium Series **41**: 95-98.

Harris, S., Morris, P., Wray, S. and Yalden, D. (1995). A review of British mammals:population estimates and conservation status of British mammals other than cetaceans. Peterborough, UK, Joint Nature Conservation Committee.

Harvell, C. D., Mitchell, C. E., Ward, J. R., Altizer, S., Dobson, A. P., Ostfeld, R. S. and Samuel, M. D. (2002). "Climate warming and disease risks for terrestrial and marine biota." Science **296**(5576): 2158-2162.

Heaton, M. P., Grosse, W. M., Kappes, S. M., Keele, J. W., Chitko-McKown, C. G., Cundiff, L. V., Braun, A., Little, D. P. and Laegreid, W. W. (2001). "Estimation of DNA sequence diversity in bovine cytokine genes." Mammalian Genome **12**(1): 32-37.

Hedrick, P. W. (2002). "Pathogen resistance and genetic variation at MHC loci." Evolution **56**(10): 1902-1908.

Hill, A. V. S. (1998). "The immunogenetics of human infectious diseases." Annual Review of Immunology **16**: 593-617.

Hill, A. V. S. (2001a). "The genomics and genetics of human infectious disease susceptibility." Annual Review of Genomics and Human Genetics **2**(1): 373-400.

Hill, A. V. S. (2001b). "Immunogenetics and genomics." The Lancet **357**(9273): 2037-2041.

Hill, A. V. S. (2006). "Aspects of Genetic Susceptibility to Human Infectious Diseases." Annual Review of Genetics **40**(1): 469-486.

Hollegaard, M. V. and Bidwell, J. L. (2006). "Cytokine gene polymorphism in human disease: On-line databases, Supplement 3." Genes and Immunity **7**(4): 269-276.

Holm, S. (1979). "A Simple Sequentially Rejective Multiple Test Procedure." Scandinavian Journal of Statistics **6**(2): 65-70.

Hori, S., Nomura, T. and Sakaguchi, S. (2003). "Control of regulatory T cell development by the transcription factor Foxp3." Science **299**(5609): 1057-1061.

Huang, H., Winter, E. E., Wang, H., Weinstock, K. G., Xing, H., Goodstadt, L., Stenson, P. D., Cooper, D. N., Smith, D., Albã , M. M., Ponting, C. P. and Fechtel, K. (2004). "Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes." Genome biology **5**(7).

Hudson, R. R. (1990). "Gene genealogies and the coalescent process." Oxford Surveys in Evolutionary Biology **7**: 1-44.

Hudson, R. R., Kreitman, M. and Aguade, M. (1987). "A Test of Neutral Molecular Evolution Based on Nucleotide Data." Genetics **116**(1): 153-159.

Huggett, J., Dheda, K., Bustin, S. and Zumla, A. (2005). "Real-time RT-PCR normalisation; strategies and considerations." Genes and Immunity **6**(4): 279-284.

Iseki, H., Takabatake, N., Ota, N., Ishigame, T., Yokoyama, N. and Igarashi, I. (2008). "*Babesia*: The protective effects of killed *Propionibacterium acnes* on the infections of two rodent *Babesia* parasites in mice." Experimental Parasitology **118**(4): 543-548.

Jackson, J. A., Friberg, I. M., Bolch, L., Lowe, A., Ralli, C., Harris, P. D., Behnke, J. M. and Bradley, J. E. (2009). "Immunomodulatory parasites and toll-like receptor-mediated tumour necrosis factor alpha responsiveness in wild mammals." BMC Biology **7**.

Jackson, J. A., Paterson, S., Birtles, R., Ralli, C., Lowe, A., Zawadzka, M., Turner, A., Hall, A., Friberg, I., Begon, M. and Bradley, J. (in review). "The

analysis of immunological networks in wild animals: a case study on immunodynamics in the field vole, *Microtus agrestis*."

Jensen, L. F., Hansen, M. M., Mensberg, K. L. D. and Loeschcke, V. (2008). "Spatially and temporally fluctuating selection at non-MHC immune genes: Evidence from TAP polymorphism in populations of brown trout (*Salmo trutta*, L.)." Heredity **100**(1): 79-91.

Jepson, A., Banya, W., Sisay-Joof, F., Hassan-King, M., Nunes, C., Bennett, S. and Whittle, H. (1997). "Quantification of the relative contribution of major histocompatibility complex (MHC) and non-MHC genes to human immune responses to foreign antigens." Infection and Immunity **65**(3): 872-876.

Johnson, J. B. and Omland, K. S. (2004). "Model selection in ecology and evolution." Trends in Ecology and Evolution **19**(2): 101-108.

Kastelein, R. A., Hunter, C. A. and Cua, D. J. (2007). "Discovery and Biology of IL-23 and IL-27: Related but Functionally Distinct Regulators of Inflammation." Annual Review of Immunology **25**(1): 221-242.

Kelso, A. (1998). "Cytokines: Principles and prospects." Immunology & Cell Biology **76**(4): 300-317.

Keymer, A. E., Tarlton, A. B., Hiorns, R. W., Lawrence, C. E. and Pritchard, D. I. (1990). "Immunogenetic correlates of susceptibility to infection with Heligmosomoides polygyrus in outbred mice." Parasitology **101**(1): 69-73.

Kimbrell, D. A. and Beutler, B. (2001). "The evolution and genetics of innate immunity." Nature Reviews Genetics **2**(4): 256-267.

Kimura, M. (1969). "The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations." Genetics **61**(4): 893-903

Kjemtrup, A. M. and Conrad, P. A. (2000). "Human babesiosis: An emerging tick-borne disease." International Journal for Parasitology **30**(12-13): 1323-1337.

Kumar, P., Henikoff, S. and Ng, P. C. (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nature Protocols **4**(7): 1073-1082.

Kunz, M. and Ibrahim, S. M. (2009). "Cytokines and cytokine profiles in human autoimmune diseases and animal models of autoimmunity." Mediators of Inflammation **2009**.

Kurtz, J., Kalbe, M., Aeschlimann, P. B., Häberli, M. A., Wegner, K. M., Reusch, T. B. H. and Milinski, M. (2004). "Major histocompatibility complex diversity influences parasite resistance and innate immunity in sticklebacks." Proceedings of the Royal Society B: Biological Sciences **271**(1535): 197-204.

Lambin, X., Elston, D. A., Petty, S. J. and MacKinnon, J. L. (1998). "Spatial asynchrony and periodic travelling waves in cyclic populations of field voles." Proceedings of the Royal Society B: Biological Sciences **265**(1405): 1491-1496.

Lambin, X., Petty, S. J. and Mackinnon, J. L. (2000). "Cyclic dynamics in field vole populations and generalist predation." Journal of Animal Ecology **69**(1): 106-118.

Letterio, J. J. and Roberts, A. B. (1998). Regulation of immune responses by TGF-β. Annual Review of Immunology. **16:** 137-161.

Levi-Acobas, F., Mars, L. T., Orth, A., Bureau, J. F. and Bonhomme, F. (2009). "Adaptive evolution of interferon-γ in Glire lineage and evidence for a recent selective sweep in *Mus. m. domesticus*." Genes and Immunity **10**(4): 297-308.

Lewontin, R. C. and Krakauer, J. (1973). "Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms." Genetics **74**(1): 175-195.

Librado, P. and Rozas, J. (2009). "DnaSP v5: A software for comprehensive analysis of DNA polymorphism data." Bioinformatics **25**(11): 1451-1452.

Livak, K. J. and Schmittgen, T. D. (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method." Methods **25**(4): 402-408.

Luikart, G., England, P. R., Tallmon, D., Jordan, S. and Taberlet, P. (2003). "The power and promise of population genomics: From genotyping to genome typing." Nature Reviews Genetics **4**(12): 981-994.

Mackey, K., Steinkamp, A. and Chomczynski, P. (1998). "DNA extraction from small blood volumes and the processing of cellulose blood cards for use in polymerase chain reaction." Applied Biochemistry and Biotechnology - Part B Molecular Biotechnology **9**(1): 1-5.

Malo, D., Vogan, K., Vidal, S., Hu, J., Cellier, M., Schurr, E., Fuks, A., Bumstead, N., Morgan, K. and Gros, P. (1994). "Haplotype Mapping and Sequence Analysis of the Mouse Nramp Gene Predict Susceptibility to Infection with Intracellular Parasites." Genomics **23**(1): 51-61.

Maynard Smith, J. and Haigh, J. (1974). "The hitch hiking effect of a favourable gene." Genetical Research **23**(1): 23-35.

McDonald, J. H. and Kreitman, M. (1991). "Adaptive protein evolution at the Adh locus in Drosophila." Nature **351**(6328): 652-654.

Meer-Scherrer, L., Adelson, M., Mordechai, E., Lottaz, B. and Tilton, R. (2004). "*Babesia microti* infection in Europe." Current Microbiology **48**(6): 435-437.

Metzker, M. L. (2010). "Sequencing technologies - the next generation." Nature Reviews Genetics **11**(1): 31-46.

Morin, P. A., Luikart, G., Wayne, R. K. and the, S. N. P. w. g. (2004). "SNPs in ecology, evolution and conservation." Trends in Ecology & Evolution **19**(4): 208-216.

Mosmann, T. R., Cherwinski, H. and Bond, M. W. (1986). "Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins." Journal of Immunology **136**(7): 2348-2357.

Mullen, A. C., High, F. A., Hutchins, A. S., Lee, H. W., Villarino, A. V., Livingston, D. M., Kung, A. L., Cereb, N., Yao, T. P., Yang, S. Y. and Reiner, S. L. (2001). "Role of T-bet in commitment of Th1 cells before IL-12-dependent selection." Science **292**(5523): 1907-1910.

Musso, T., Badolato, R., Ravarino, D., Stornello, S., Panzanelli, P., Merlino, C., Savoia, D., Cavallo, R., Negro Ponzi, A. and Zucca, M. (2001). "Interaction of *Bartonella henselae* with the murine macrophage cell line J774: Infection and proinflammatory response." Infection and Immunity **69**(10): 5974-5980.

Nachman, M. W. (2006). Detecting Selection at the Molecular Level. Evolutionary Genetics: Concepts and Case Studies. C. W. Fox and J. B. Wolf. New York, Oxford University Press**:** 103-118.

Nakajima, R., Tsuji, M., Oda, K., Zamoto-Niikura, A., Wei, Q., Kawabuchi-Kurata, T., Nishida, A. and Ishihara, C. (2009). "*Babesia microti*-group parasites compared phylogenetically by complete sequencing of the CCTη· gene in 36 isolates." Journal of Veterinary Medical Science **71**(1): 55-68.

Nei, M. (1987). Molecular Evolutionary Genetics. New York, Columbia University Press.

Nielsen, R. (2005). Molecular signatures of natural selection. Annual Review of Genetics. **39:** 197-218.

Nielsen, R. and Signorovitch, J. (2003). "Correcting for ascertainment biases when analyzing SNP data: Applications to the estimation of linkage disequilibrium." Theoretical Population Biology **63**(3): 245-255.

O'Donovan, M. R., Johns, S. and Wilcox, P. (1995). "The effect of PHA stimulation on lymphocyte sub-populations in whole-blood cultures." Mutagenesis **10**(4): 371-374.

O'Shea, J. J., Ma, A. and Lipsky, P. (2002). "Cytokines and autoimmunity." Nature Reviews Immunology **2**(1): 37-45.

Obbard, D. J., Welch, J. J., Kim, K.-W. and Jiggins, F. M. (2009). "Quantifying adaptive evolution in the *Drosophila* immune system." PLoS Genetics **5**(10): e1000698.

Ohta, T. (1982). "Linkage disequilibrium due to random genetic drift in finite subdivided populations." Proceedings of the National Academy of Sciences of the United States of America **79**(6 I): 1940-1944.

Oko, L., Aduddell-Swope, B., Willis, D., Hamor, R., Coons, T. A., Hjelle, B. and Schountz, T. (2006). "Profiling helper T cell subset gene expression in deer mice." BMC Immunology **7**.

Oliver, M. K., Telfer, S. and Piertney, S. B. (2009). "Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*)." Proceedings of the Royal Society B: Biological Sciences **276**(1659): 1119-1128.

Ollier, W. E. R. (2004). "Cytokine genes and disease susceptibility." Cytokine **28**(4-5): 174-178.

Oppmann, B., Lesley, R., Blom, B., Timans, J. C., Xu, Y., Hunte, B., Vega, F., Yu, N., Wang, J., Singh, K., Zonin, F., Vaisberg, E., Churakova, T., Liu, M.-r., Gorman, D., Wagner, J., Zurawski, S., Liu, Y.-J., Abrams, J. S., Moore, K. W., Rennick, D., de Waal-Malefyt, R., Hannum, C., Bazan, J. F. and Kastelein, R. A. (2000). "Novel p19 Protein Engages IL-12p40 to Form a Cytokine, IL-23, with Biological Activities Similar as Well as Distinct from IL-12." Immunity **13**(5): 715-725.

Ozaki, K. and Leonard, W. J. (2002). "Cytokine and cytokine receptor pleiotropy and redundancy." Journal of Biological Chemistry **277**(33): 29355-29358.

Paterson, S., Wilson, K. and Pemberton, J. M. (1998). "Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L.)." Proceedings of the National Academy of Sciences of the United States of America **95**(7): 3714-3719.

Paul, W. E. and Zhu, J. (2010). "How are Th2-type immune responses initiated and amplified?" Nature Reviews Immunology **10**(4): 225-235.

Paun, A., Reinert, J. T., Jiang, Z., Medin, C., Balkhi, M. Y., Fitzgerald, K. A. and Pitha, P. M. (2008). "Functional characterization of murine interferon regulatory factor 5 (IRF-5) and its role in the innate antiviral response." Journal of Biological Chemistry **283**(21): 14295-14308.

Piertney, S. B. and Oliver, M. K. (2006). "The evolutionary ecology of the major histocompatibility complex." Heredity **96**(1): 7-21.

Piertney, S. B. and Webster, L. M. I. (2010). "Characterising functionally important and ecologically meaningful genetic diversity using a candidate gene approach." Genetica **138**(4): 419-432.

Pompanon, F., Bonin, A., Bellemain, E. and Taberlet, P. (2005). "Genotyping errors: Causes, consequences and solutions." Nature Reviews Genetics **6**(11): 847-859.

Popa, C., Abdollahi-Roodsaz, S., Joosten, L. A. B., Takahashi, N., Sprong, T., Matera, G., Liberto, M. C., Foca, A., Van Deuren, M., Kullberg, B. J., Van Den Berg, W. B., Van Der Meer, J. W. M. and Netea, M. G. (2007). "*Bartonella quintana* lipopolysaccharide is a natural antagonist of toll-like receptor 4." Infection and Immunity **75**(10): 4831-4837.

R Development Core Team (2009). R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.

Ramensky, V., Bork, P. and Sunyaev, S. (2002). "Human non-synonymous SNPs: Server and survey." Nucleic Acids Research **30**(17): 3894-3900.

Raymond, M. and Rousset, F. (1995). "GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism." Journal of Heredity **86**(3): 248-249.

Reed, S. G. (1999). "TGF-β in infections and infectious diseases." Microbes and Infection **1**(15): 1313-1325.

Resto-Ruiz, S. I., Schmiederer, M., Sweger, D., Newton, C., Klein, T. W., Friedman, H. and Anderson, B. E. (2002). "Induction of a potential paracrine angiogenic loop between human THP-1 macrophages and human microvascular endothelial cells during *Bartonella henselae* infection." Infection and Immunity **70**(8): 4564-4570.

Reusch, T. B. H., Ha?berli, M. A., Aeschlimann, P. B. and Milinski, M. (2001). "Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism." Nature **414**(6861): 300-302.

Rice, W. R. (1989). "Analyzing tables of statistical tests." Evolution **43**(1): 223-225.

Romani, L., Puccetti, P. and Bistoni, F. (1997). "Interleukin-12 in infectious diseases." Clinical Microbiology Reviews **10**(4): 611-636.

Rousset, F. (2008). "GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux." Molecular Ecology Resources **8**(1): 103-106.

Sacks, B. N. and Louie, S. (2008). "Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae." Molecular Ecology Resources **8**(1): 35-49.

Sackton, T. B., Lazzaro, B. P., Schlenke, T. A., Evans, J. D., Hultmark, D. and Clark, A. G. (2007). "Dynamic evolution of the innate immune system in Drosophila." Nature Genetics **39**(12): 1461-1468.

Schlenke, T. A. and Begun, D. J. (2003). "Natural selection drives Drosophila immune system evolution." Genetics **164**(4): 1471-1480.

Schoeler, G. B., Manweiler, S. A. and Wikel, S. K. (1999). "*Ixodes scapularis*: Effects of repeated infestations with pathogen-free nymphs on macrophage and T lymphocyte cytokine responses of BALB/c and C3H/HeN mice." Experimental Parasitology **92**(4): 239-248.

Schröder, N. W. J. and Schumann, R. R. (2005). "Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease." The Lancet Infectious Diseases **5**(3): 156-164.

Shoda, L. K. M., Palmer, G. H., Florin-Christensen, J., Florin-Christensen, M., Godson, D. L. and Brown, W. C. (2000). "*Babesia bovis*-stimulated macrophages express interleukin-1β, interleukin-12, tumor necrosis factor alpha, and nitric oxide and inhibit parasite replication *in vitro*." Infection and Immunity **68**(9): 5139-5145.

Sims, J. E. and Smith, D. E. (2010). "The IL-1 family: Regulators of immunity." Nature Reviews Immunology **10**(2): 89-102.

Slate, J., Gratten, J., Beraldi, D., Stapley, J., Hale, M. and Pemberton, J. (2009). "Gene mapping in the wild with SNPs: guidelines and future directions." Genetica **136**(1): 97-107.

Slate, J., Santure, A. W., Feulner, P. G. D., Brown, E. A., Ball, A. D., Johnston, S. E. and Gratten, J. (2010). "Genome mapping in intensively studied wild vertebrate populations." Trends in Genetics **26**(6): 275-284.

Slev, P. R. and Potts, W. K. (2002). "Disease consequences of pathogen adaptation." Current Opinion in Immunology **14**(5): 609-614.

Smith, A., Telfer, S., Burthe, S., Bennett, M. and Begon, M. (2005). "Trypanosomes, fleas and field voles: Ecological dynamics of a host-vector-parasite interaction." Parasitology **131**(3): 355-365.

Smith, A. J. P. and Humphries, S. E. (2008). "Cytokine and cytokine receptor gene polymorphisms and their functionality." Cytokine & Growth Factor Reviews **20**(1).

Smith, A. J. P. and Humphries, S. E. (2009). "Cytokine and cytokine receptor gene polymorphisms and their functionality." Cytokine & Growth Factor Reviews **20**(1): 43-59.

Smith, K. F., Acevedo-Whitehouse, K. and Pedersen, A. B. (2009). "The role of infectious diseases in biological conservation." Animal Conservation **12**(1): 1-12.

Sommer, S. (2005). "The importance of immune gene variability (MHC) in evolutionary ecology and conservation." Frontiers in Zoology **2**.

Stearns, S. C. (1999). Evolution in Health and Disease. Oxford, Oxford University Press, Oxford.

Steinke, J. W. and Borish, L. (2006). "3. Cytokines and chemokines." Journal of Allergy and Clinical Immunology **117**(2 SUPPL. 2).

Stephens, M. and Scheet, P. (2005). "Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation." American Journal of Human Genetics **76**(3): 449-462.

Stephens, M., Smith, N. J. and Donnelly, P. (2001). "A new statistical method for haplotype reconstruction from population data." American Journal of Human Genetics **68**(4): 978-989.

Steppan, S., Adkins, R. and Anderson, J. (2004). "Phylogeny and Divergence-Date Estimates of Rapid Radiations in Muroid Rodents Based on Multiple Nuclear Genes." Systematic Biology **53**: 533-553.

Tajima, F. (1989). "Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." Genetics **123**(3): 585-595.

Takahata, N. and Nei, M. (1990). "Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci." Genetics **124**(4): 967-978.

Takeda, K., Kaisho, T. and Akira, S. (2003). Toll-like receptors. Annual Review of Immunology. **21:** 335-376.

Tautz, D., Ellegren, H. and Weigel, D. (2010). "Next Generation Molecular Ecology." Molecular Ecology **19**: 1-3.

Telfer, S., Begon, M., Bennett, M., Bown, K. J., Burthe, S., Lambin, X., Telford, G. and Birtles, R. (2007a). "Contrasting dynamics of *Bartonella* spp. in cyclic field vole populations: The impact of vector and host dynamics." Parasitology **134**(3): 413-425.

Telfer, S., Birtles, R., Bennett, M., Lambin, X., Paterson, S. and Begon, M. (2008). "Parasite interactions in natural populations: insights from longitudinal data." Parasitology **135**(Special Issue 07): 767-781.

Telfer, S., Bown, K. J., Sekules, R., Begon, M., Hayden, T. and Birtles, R. (2005). "Disruption of a host-parasite system following the introduction of an exotic host species." Parasitology **130**(6): 661-668.

Telfer, S., Clough, H. E., Birtles, R. J., Bennett, M., Carslake, D., Helyar, S. and Begon, M. (2007b). "Ecological differences and coexistence in a guild of microparasites: *Bartonella* in wild rodents." Ecology **88**(7): 1841-1849.

Telfer, S., Lambin, X., Birtles, R., Beldomenico, P., Burthe, S., Paterson, S. and Begon, M. (in press). "Species interactions in a parasite community drive infection risk in a wildlife population." Science.

Tennessen, J. A. and Blouin, M. S. (2008). "Balancing selection at a frog antimicrobial peptide locus: Fluctuating immune effector alleles?" Molecular Biology and Evolution **25**(12): 2669-2680.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Research **22**(22): 4673-4680.

Tonteri, A., Vasemägi, A., Lumme, J. and Primmer, C. R. (2010). "Beyond MHC: Signals of elevated selection pressure on Atlantic salmon (*Salmo salar*) immune-relevant loci." Molecular Ecology **19**(7): 1273-1282.

Trowsdale, J. and Parham, P. (2004). "Mini-review: Defense strategies and immunity-related genes." European Journal of Immunology **34**(1): 7-17.

Uilenberg, G. (2006). "Babesia-A historical overview." Veterinary Parasitology **138**(1-2): 3-10.

van Bers, N. E. M., van Oers, K., Kerstens, H. H. D., Dibbits, B. W., Crooijmans, R. P. M. A., Visser, M. E. and Groenen, M. A. M. (2010). "Genome-wide SNP detection in the great tit *Parus major* using high throughput sequencing." Molecular Ecology **19**(s1): 89-99.

Van Deventer, S. J. H. (2000). "Cytokine and cytokine receptor polymorphisms in infectious disease." Intensive Care Medicine, Supplement **26**(1).

Vasemägi, A. and Primmer, C. R. (2005). "Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies." Molecular Ecology **14**(12): 3623-3642.

Venables, W. N. and Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth edition. New York, Springer.

Vera, J. C., Wheat, C. W., Fescemyer, H. W., Frilander, M. J., Crawford, D. L., Hanski, I. and Marden, J. H. (2008). "Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing." Molecular Ecology **17**(7): 1636-1647.

Vidal, S. M. (1993). "Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg." Trends in Genetics **9**(8): 264-264.

Watterson, G. A. (1975). "On the number of segregating sites in genetical models without recombination." Theoretical Population Biology **7**(2): 256-276.

Watterson, G. A. (1978). "The homozygosity test of neutrality." Genetics **88**(2): 405-417.

Weaver, C. T., Harrington, L. E., Mangan, P. R., Gavrieli, M. and Murphy, K. M. (2006). "Th17: An Effector CD4 T Cell Lineage with Regulatory T Cell Ties." Immunity **24**(6): 677-688.

Weir, B. S. (1979). "Inferences about Linkage Disequilibrium." Biometrics **35**(1): 235-254.

Weir, B. S. (1996). Genetic Data Analysis II. Sunderland, MA., Sinaur Associates, Inc.

Weir, B. S. and Cockerham, C. C. (1984). "Estimating F-Statistics for the Analysis of Population Structure." Evolution **38**(6): 1358-1370.

Wikel, S. K. (1999). "Modulation of the Host Immune System by Ectoparasitic Arthropods." BioScience **49**(4): 311-320.

Worley, K., Carey, J., Veitch, A. and Coltman, D. W. (2006). "Detecting the signature of selection on immune genes in highly structured populations of wild sheep (*Ovis dalli*)." Molecular Ecology **15**(3): 623-637.

Wright, S. (1951). "The genetical structure of populations." Annals of Eugenics **15**: 323-254.

Yang, Y., Li, S. S., Chien, J. W., Andriesen, J. and Zhao, L. P. (2008). "A systematic search for SNPs/haplotypes associated with disease phenotypes using a haplotype-based stepwise procedure." BMC Genetics **9**.

Yoshimoto, T., Okamura, H., Tagawa, Y. I., Iwakura, Y. and Nakanishi, K. (1997). "Interleukin 18 together with interleukin 12 inhibits IgE production by induction of interferon-γ production from activated B cells." Proceedings of the National Academy of Sciences of the United States of America **94**(8): 3948-3953.

Zhu, J., Yamane, H., Cote-Sierra, J., Guo, L. and Paul, W. E. (2006). "GATA-3 promotes Th2 responses through three different mechanisms: Induction of Th2 cytokine production, selective growth of Th2 cells and inhibition of Th1 cell-specific factors." Cell Research **16**(1): 3-10.

Appendix 1 - Genotyping data

Appendix Table 1.1 SNP frequency data for BLB, SQC and combined sites

Gene	SNP	Site	*N*	MAF	H_O	H_E	HWE *p*	F_{IS}
Il1b	Il1b 243 G/A	Both	571	0.16	0.26	0.26	0.853	0.000
		BLB	268	0.19	0.31	0.30	0.839	-0.026
		SQC	303	0.13	0.22	0.22	0.607	0.031
	Il1b 253 A/G	Both	568	0.32	0.43	0.44	0.745	0.025
		BLB	268	0.29	0.39	0.42	0.378	0.056
		SQC	300	0.35	0.46	0.46	1.000	0.000
	Il1b 324 C/T	Both	568	0.21	0.33	0.33	0.276	0.001
		BLB	268	0.20	0.34	0.32	0.248	-0.076
		SQC	300	0.22	0.32	0.35	0.314	0.065
Il2	Il2 381 A/T	Both	562	0.14	0.25	0.25	0.108	-0.039
		BLB	265	0.12	0.24	0.21	**0.033**	-0.133
		SQC	297	0.17	0.27	0.28	0.675	0.024
	Il2 408 C/G	Both	567	0.29	0.41	0.41	0.188	0.005
		BLB	267	0.32	0.47	0.44	0.261	-0.074
		SQC	300	0.26	0.35	0.39	0.176	0.083
Il12b	Il12b 278 G/C	Both	570	0.05	0.10	0.10	0.920	0.012
		BLB	270	0.07	0.13	0.13	0.628	0.039
		SQC	300	0.04	0.07	0.07	1.000	-0.035
	Il12b 704 C/T	Both	569	0.02	0.03	0.03	1.000	0.082
		BLB	270	0.03	0.06	0.06	0.254	0.082
		SQC	299	0.00	-	-	-	-
Slc11a1	Slc11a1 537 C/G	Both	568	0.26	0.39	0.39	0.854	-0.022
		BLB	268	0.30	0.42	0.42	1.000	-0.001
		SQC	300	0.23	0.37	0.35	0.511	-0.044
	Slc11a1 714 G/A	Both	571	0.28	0.41	0.40	0.798	-0.022
		BLB	269	0.32	0.43	0.43	1.000	0.004
		SQC	302	0.24	0.39	0.37	0.436	-0.050
Tgfb1	Tgfb1 442 C/A	Both	565	0.25	0.37	0.38	0.071	0.018
		BLB	266	0.24	0.40	0.37	0.179	-0.089
		SQC	299	0.26	0.35	0.39	0.074	0.107
Tlr2	Tlr2 1383 G/A	Both	576	0.06	0.10	0.11	0.516	0.041
		BLB	267	0.06	0.11	0.11	1.000	0.005
		SQC	309	0.05	0.09	0.10	0.196	0.078
	Tlr2 1648 G/A	Both	564	0.10	0.18	0.18	0.968	0.008
		BLB	265	0.09	0.17	0.17	1.000	-0.010
		SQC	299	0.11	0.19	0.19	0.759	0.022
	Tlr2 1706 G/A	Both	571	0.001	-	-	-	0.000
		BLB	272	0.002	0.00	0.00	-	-
		SQC	299	0.00	-	-	-	-
	Tlr2 1944 T/C	Both	565	0.30	0.39	0.42	**0.004**	0.070
		BLB	269	0.29	0.33	0.41	**0.001**	0.207
		SQC	296	0.31	0.45	0.43	0.421	-0.049

Appendix Table 1.1 continued

Gene	SNP	Site	*N*	MAF	H_O	H_E	HWE *p*	F_{IS}
Tlr4	Tlr4 1663 A/G	Both	561	0.06	0.10	0.10	0.163	0.079
		BLB	267	0.06	0.10	0.12	0.078	0.122
		SQC	294	0.05	0.09	0.09	0.491	0.027
	Tlr4 1848 G/T	Both	559	0.02	0.04	0.04	1.000	-0.024
		BLB	262	0.03	0.06	0.06	1.000	-0.030
		SQC	297	0.02	0.03	0.03	1.000	-0.014
	Tlr4 2037 C/A	Both	563	0.05	0.10	0.10	0.908	-0.018
		BLB	267	0.06	0.11	0.11	0.603	0.009
		SQC	296	0.05	0.09	0.09	1.000	-0.048
Tnf	Tnf 210 T/C	Both	570	0.37	0.46	0.47	0.098	0.019
		BLB	269	0.37	0.50	0.47	0.240	-0.075
		SQC	301	0.37	0.42	0.47	0.083	0.103
			Average	0.16	0.25	0.25	0.530	0.012

Appendix Table 1.2 SNP frequency data for Kielder, Kershope, Redesdale and combined sites

Gene	SNP	Site	*N*	MAF	H_O	H_E	HWE *p*	F_{IS}
Il1b	Il1b 243 G/A	All	154	0.17	0.23	0.28	0.08	0.206
		Kielder	49	0.20	0.24	0.33	0.09	0.256
		Kershope	48	0.13	0.21	0.22	0.54	0.058
		Redesdale	57	0.18	0.23	0.30	0.08	0.250
	Il1b 253 A/G	All	155	0.36	0.45	0.46	0.92	0.048
		Kielder	50	0.38	0.40	0.48	0.37	0.161
		Kershope	48	0.35	0.46	0.46	1.00	0.009
		Redesdale	57	0.36	0.47	0.46	1.00	-0.020
	Il1b 324 C/T	All	153	0.14	0.23	0.24	0.49	0.046
		Kielder	50	0.13	0.26	0.23	0.58	-0.140
		Kershope	47	0.09	0.13	0.16	0.28	0.191
		Redesdale	56	0.20	0.29	0.32	0.42	0.104
Il2	Il2 381 A/T	All	154	0.16	0.27	0.27	1.00	0.004
		Kielder	50	0.19	0.30	0.31	1.00	0.035
		Kershope	47	0.14	0.23	0.24	1.00	0.029
		Redesdale	57	0.16	0.28	0.27	1.00	-0.047
	Il2 408 C/G	All	155	0.20	0.34	0.32	0.39	-0.053
		Kielder	50	0.25	0.42	0.38	0.70	-0.110
		Kershope	48	0.19	0.25	0.31	0.33	0.190
		Redesdale	57	0.18	0.35	0.29	0.18	-0.204
Il12b	Il12b_278	All	155	0.08	0.16	0.15	1.00	-0.092
		Kielder	50	0.04	0.08	0.08	1.00	-0.032
		Kershope	48	0.12	0.23	0.21	1.00	-0.119
		Redesdale	57	0.09	0.18	0.16	1.00	-0.087
	Il12b_704	All	154	0.00	-	-	-	
		Kielder	50	0.00	-	-	-	
		Kershope	48	0.00	-	-	-	
		Redesdale	56	0.00	-	-	-	
Slc11a1	Slc11a1 537 C/G	All	153	0.27	0.36	0.40	0.44	0.099
		Kielder	50	0.27	0.42	0.40	1.00	-0.055
		Kershope	48	0.29	0.33	0.42	0.18	0.203
		Redesdale	55	0.26	0.33	0.38	0.30	0.147
	Slc11a1 714 G/A	All	153	0.29	0.38	0.42	0.23	0.097
		Kielder	49	0.30	0.47	0.42	0.50	-0.116
		Kershope	48	0.29	0.33	0.42	0.18	0.203
		Redesdale	56	0.30	0.34	0.42	0.20	0.192
Tgfb1	Tgfb1 442 C/A	All	154	0.24	0.32	0.36	0.51	0.127
		Kielder	49	0.20	0.29	0.33	0.39	0.131
		Kershope	48	0.26	0.31	0.39	0.25	0.199
		Redesdale	57	0.25	0.35	0.37	0.72	0.062

Appendix Table 1.2 continued

Gene	SNP	Site	*N*	MAF	H_O	H_E	HWE *p*	F_{IS}
Tlr2	Tlr2 1383 G/A	All	155	0.06	0.13	0.12	1.00	-0.080
		Kielder	50	0.05	0.10	0.10	1.00	-0.043
		Kershope	48	0.12	0.23	0.21	1.00	-0.119
		Redesdale	57	0.04	0.07	0.07	1.00	-0.028
	Tlr2 1648 G/A	All	153	0.04	0.08	0.08	1.00	-0.034
		Kielder	49	0.04	0.08	0.08	1.00	-0.032
		Kershope	48	0.05	0.10	0.10	1.00	-0.044
		Redesdale	56	0.03	0.05	0.05	1.00	-0.019
	Tlr2 1706 G/A	All	155	0.02	0.05	0.04	1.00	-0.021
		Kielder	50	0.01	0.02	0.02	1.00	-
		Kershope	48	0.04	0.08	0.08	1.00	-0.033
		Redesdale	57	0.02	0.04	0.03	1.00	-0.009
	Tlr2 1944 T/C	All	152	0.34	0.39	0.45	**0.03**	0.117
		Kielder	49	0.40	0.39	0.48	0.23	0.201
		Kershope	46	0.32	0.54	0.44	0.17	-0.249
		Redesdale	57	0.30	0.28	0.42	**0.02**	0.337
Tlr4	Tlr4 1663 A/G	All	154	0.06	0.11	0.10	1.00	-0.069
		Kielder	50	0.01	0.02	0.02	1.00	-
		Kershope	48	0.08	0.17	0.15	1.00	-0.081
		Redesdale	56	0.07	0.14	0.13	1.00	-0.068
	Tlr4 1848 G/T	All	153	0.02	0.13	0.12	1.00	-0.034
		Kielder	50	0.00	-	-	-	-
		Kershope	48	0.05	0.10	0.10	1.00	-0.044
		Redesdale	55	0.02	0.04	0.04	1.00	-0.009
	Tlr4 2037 C/A	All	155	0.05	0.11	0.10	1.00	-0.069
		Kielder	50	0.01	0.02	0.02	1.00	-
		Kershope	48	0.08	0.17	0.15	1.00	-0.081
		Redesdale	57	0.07	0.14	0.13	1.00	-0.067
Tnf	Tnf 210 T/C	All	155	0.37	0.46	0.47	0.46	0.027
		Kielder	50	0.35	0.54	0.46	0.35	-0.177
		Kershope	48	0.38	0.17	0.15	1.00	0.033
		Redesdale	57	0.39	0.39	0.48	0.17	0.194
			Average	0.17	0.25	0.26	0.69	0.052

Appendix Table 1.3 Observed haplotype frequencies and those expected under a model of neutrality, as predicted by Ewens' (1972) sampling formula.

Site		BLB			SQC				All			
Gene	Haplotype	Count	Obs. Frequency	Exp. Frequency	Haplotype	Count	Obs. Frequency	Exp. Frequency	Haplotype	Count	Obs. Frequency	Exp. Frequency
Il1b	GAC	168	0.36	0.69	GGC	236	0.36	0.70	GAC	362	0.32	0.72
	GGC	124	0.27	0.21	GAC	194	0.29	0.20	GGC	360	0.32	0.20
	GAT	106	0.23	0.07	GAT	141	0.21	0.07	GAT	247	0.22	0.06
	AAC	66	0.14	0.02	AAC	85	0.13	0.02	AAC	151	0.13	0.02
	GGT	2	0.00	0.01	GGT	4	0.01	0.01	GGT	6	0.01	0.00
Il2	AC	248	0.53	0.82	AC	357	0.55	0.82	AC	605	0.54	0.83
	AG	175	0.37	0.16	AG	206	0.32	0.15	AG	381	0.34	0.14
	TC	45	0.10	0.03	TC	81	0.13	0.03	TC	126	0.11	0.02
Il12b	GC	429	0.90	0.82	GC	636	0.97	0.90	GC	1065	0.94	0.83
	CC	32	0.03	0.16	CC	17	0.03	0.90	CC	49	0.04	0.14
	GT	14	0.07	0.03					GT	14	0.01	0.02
Slc11a1	CG	313	0.65	0.75	CG	488	0.74	0.82	CG	801	0.70	0.77
	GA	153	0.32	0.19	GA	152	0.23	0.15	GA	305	0.27	0.18
	GG	8	0.02	0.05	CA	18	0.03	0.03	CA	24	0.02	0.04
	CA	6	0.01	0.01	-	-	-	-	GG	8	0.01	0.01
Tgfb1	C	297	0.65	0.90	C	498	0.77	0.90	C	795	0.72	0.91
	A	157	0.35	0.10	A	152	0.23	0.10	A	309	0.28	0.09
Tlr2	GGGC	260	0.56	0.69	GGGC	339	0.51	0.76	GGGC	599	0.53	0.72
	GGGT	129	0.28	0.21	GGGT	214	0.32	0.18	GGGT	343	0.30	0.20
	GAGC	52	0.11	0.07	GAGC	75	0.11	0.05	GAGC	127	0.11	0.06
	AGGC	24	0.05	0.02	AGGC	32	0.05	0.01	AGGC	56	0.05	0.02
	GAGT	3	0.01	0.01	-	-	-	-	GAGT	3	0.00	0.00
Tlr4	AGC	421	0.92	0.69	AGC	617	0.95	0.82	AGC	1038	0.94	0.72
	GTA	21	0.05	0.21	GGA	25	0.04	0.15	GTA	34	0.03	0.20
	GGA	9	0.02	0.07	GTA	6	0.01	0.03	GGA	27	0.02	0.06
	GGC	4	0.01	0.02	-	-	-	-	GGC	4	0.00	0.02
	AGA	3	0.01	0.01	-	-	-	-	AGA	3	0.00	0.00
Tnf	C	282	0.60	0.90	C	404	0.62	0.90	C	686	0.61	0.91
	A	188	0.40	0.10	A	246	0.38	0.10	A	434	0.39	0.09

Appendix Table 1.4 Haplotype diversity and summary of Ewens-Watterson tests of neutrality for the Kielder, Kershope, and Redesdale sites. Tests were one-sided; significance assumed where $0.05 > p > 0.95$. Significant departures from neutral expectations are in bold.

Gene	Site	2*n*	N_h [a]	H_d [b]	Obs. *F* [c]	Exp. *F* [d]	*p*-value
Il1b	Kielder	100	5	0.728	0.280	0.488	**0.044**
	Kershope	96	4	0.668	0.339	0.566	0.066
	Redesdale	114	4	0.737	0.270	0.576	**0.003**
	All	310	5	0.717	0.285	0.555	**0.024**
Il2	Kielder	100	3	0.594	0.412	0.673	0.064
	Kershope	96	3	0.493	0.512	0.670	0.261
	Redesdale	114	3	0.504	0.500	0.680	0.215
	All	310	3	0.531	0.471	0.723	0.090
Il12b	Kielder	100	2	0.078	0.923	0.809	0.637
	Kershope	96	2	0.205	0.797	0.807	0.408
	Redesdale	114	2	0.162	0.840	0.815	0.448
	All	310	2	0.149	0.852	0.845	0.382
Slc11a1	Kielder	100	3	0.427	0.577	0.671	0.380
	Kershope	96	2	0.418	0.586	0.809	0.176
	Redesdale	114	3	0.433	0.571	0.677	0.361
	All	310	3	0.424	0.577	0.725	0.294
Tgfb1	Kielder	100	2	0.323	0.680	0.810	0.268
	Kershope	96	2	0.389	0.615	0.806	0.210
	Redesdale	114	2	0.374	0.629	0.814	0.214
	All	310	2	0.361	0.640	0.840	0.191
Tlr2	Kielder	100	5	0.590	0.416	0.490	0.389
	Kershope	96	5	0.650	0.357	0.486	0.235
	Redesdale	114	5	0.513	0.491	0.498	0.566
	All	310	5	0.584	0.418	0.554	0.265
Tlr4	Kielder	100	2	0.020	0.980	0.808	**1.000**
	Kershope	96	3	0.158	0.843	0.669	0.756
	Redesdale	114	3	0.134	0.868	0.680	0.780
	All	310	3	0.106	0.895	0.723	0.722
Tnf	Kielder	100	2	0.496	0.545	0.809	0.122
	Kershope	96	2	0.474	0.531	0.809	0.103
	Redesdale	114	2	0.478	0.526	0.812	0.093
	All	310	2	0.468	0.533	0.843	0.079
		Average (SD)	3.1 (1.2)	0.42 (0.21)	0.58 (0.21)	0.72 (0.12)	0.25

[a] Number of haplotypes

[b] Haplotype diversity

[c] Observed values of Watterson's test statistic *F*

[d] Expected values of *F* under neutrality given the population size and number of haplotypes (Ewens 1972)

Appendix 2 - Minimal non-genetic models

Appendix Table 2.1 GLMs describing non-genetic factors associated with variation in the probability of infection by various parasites

Term	Coefficient	s.e	*z*-value	*p*-value
Nematode infection				
Intercept	-0.70	0.40	-1.72	0.085
Site (SQC)	-0.71	0.31	-2.30	0.021
Season (summer 2008)	-0.47	0.42	-1.12	0.265
Season (autumn 2008)	-1.34	0.42	-3.22	0.001
Season (winter 2008)	-1.30	0.44	-2.98	0.003
Season (spring 2009)	-1.53	0.82	-1.87	0.061
Sex (male)	0.82	0.36	2.29	0.022
Cestode infection				
Intercept	2.83	1.61	1.75	0.080
Season (summer 2008)	-3.14	2.15	-1.46	0.144
Season (autumn 2008)	-4.89	1.79	-2.73	0.006
Season (winter 2008)	1.14	2.31	0.49	0.623
Season (spring 2009)	-0.08	3.07	-0.03	0.979
Body weight	0.06	0.04	1.66	0.097
Eye lens weight	-999.34	255.92	-3.91	<0.0001
Sex (male)	0.88	0.33	2.62	0.009
Season (summer 2008) × body weight	0.18	0.08	2.11	0.035
Season (autumn 2008) × body weight	0.16	0.07	2.19	0.028
Season (winter 2008) × body weight	-0.20	0.12	-1.76	0.078
Season (spring 2009) × body weight	-0.07	0.15	-0.48	0.631
Flea infection				
Intercept	-0.27	1.10	-0.25	0.807
Site (SQC)	-1.14	0.60	-1.89	0.058
Season (summer 2008)	-3.10	1.91	-1.62	0.105
Season (autumn 2008)	-0.83	1.66	-0.50	0.619
Season (winter 2008)	3.34	2.39	1.40	0.162
Season (spring 2009)	5.04	3.03	1.66	0.096
Body weight	0.04	0.04	1.14	0.256
Site (SQC) × Season (summer 2008)	1.59	0.97	1.63	0.103
Site (SQC) × Season (autumn 2008)	0.26	0.75	0.35	0.724
Site (SQC) × Season (winter 2008)	2.82	0.85	3.30	0.001
Site (SQC) × Season (spring 2009)	-	-	-	-
Season (summer 2008) × body weight	0.12	0.07	1.77	0.077
Season (autumn 2008) × body weight	0.05	0.07	0.66	0.507
Season (winter 2008) × body weight	-0.31	0.13	-2.40	0.016
Season (spring 2009) × body weight	-0.24	0.15	-1.59	0.112

Appendix Table 2.1 Continued

Term	Coefficient	s.e	*z*-value	*p*-value
Tick infection				
Intercept	2.25	2.76	0.82	0.415
Season (summer 2008)	-6.66	3.07	-2.17	0.030
Season (autumn 2008)	-5.16	3.07	-1.68	0.093
Season (winter 2008)	-11.13	4.44	-2.51	0.012
Season (spring 2009)	-18.62	754.50	-0.03	0.980
Body weight	0.09	0.03	2.93	0.003
Eye lens weight	-1052.00	538.20	-1.95	0.051
Sex (male)	0.35	0.75	0.47	0.637
Season (summer 2008) × eye lens weight	1809.00	628.70	2.88	0.004
Season (autumn 2008) × eye lens weight	676.90	647.10	1.05	0.296
Season (winter 2008) × eye lens weight	2506.00	1153.00	2.17	0.030
Season (spring 2009) × eye lens weight	530.40	1464.00	0.36	0.717
Season (summer 2008) × sex (male)	-1.73	1.16	-1.50	0.134
Season (autumn 2008) × sex (male)	0.88	1.02	0.86	0.390
Season (winter 2008) × sex (male)	-0.59	1.21	-0.49	0.627
Season (spring 2009) × sex (male)	15.67	754.50	0.02	0.983
Babesia microti *infection*				
Intercept	-2.56	0.76	-3.35	0.001
Season (summer 2008)	-1.14	0.51	-2.23	0.026
Season (autumn 2008)	0.41	0.44	0.93	0.352
Season (winter 2008)	0.08	0.49	0.16	0.872
Season (spring 2009)	-1.05	0.83	-1.27	0.205
Body weight	0.06	0.02	2.48	0.013
Bartonella *infection*				
Intercept	-5.27	4.25	-1.24	0.215
Site (SQC)	0.92	0.36	2.55	0.011
Season (summer 2008)	5.72	4.39	1.30	0.192
Season (autumn 2008)	2.25	4.49	0.50	0.616
Season (winter 2008)	7.91	4.85	1.63	0.103
Season (spring 2009)	-1.71	5.61	-0.30	0.761
Eye lens weight	517.87	780.52	0.66	0.507
Season (summer 2008) × eye lens weight	-957.80	823.86	-1.16	0.245
Season (autumn 2008) × eye lens weight	-302.78	897.27	-0.34	0.736
Season (winter 2008) × eye lens weight	-2312.87	1144.28	-2.02	0.043

Appendix Table 2.2 GLMs describing non-genetic factors associated with variation in parasite burden per individual

Term	Coefficient	s.e	*t*-value	*p*-value
Cestode burden				
Intercept	0.40	1.32	0.31	0.760
Site (SQC)	1.27	0.82	1.55	0.123
Season (summer 2008)	2.56	0.92	2.78	0.006
Season (autumn 2008)	-1.10	1.11	-0.99	0.322
Season (winter 2008)	2.18	2.28	0.96	0.340
Season (spring 2009)	-0.06	2.57	-0.02	0.981
Body weight	0.01	0.03	0.52	0.601
Eye lens weight	-477.49	259.84	-1.84	0.067
Sex (male)	-1.61	0.79	-2.02	0.044
Site (SQC) × Season (summer 2008)	-1.98	0.45	-4.41	<0.0001
Site (SQC) × Season (autumn 2008)	-1.08	0.57	-1.89	0.060
Site (SQC) × Season (winter 2008)	-0.74	0.79	-0.95	0.345
Site (SQC) × Season (spring 2009)	-	-	-	-
Site (SQC) × body weight	0.08	0.03	2.90	0.004
Site (SQC) × eye lens weight	-555.25	180.88	-3.07	0.002
Season (summer 2008) × body weight	-0.02	0.03	-0.77	0.444
Season (autumn 2008) × body weight	0.08	0.04	2.09	0.037
Season (winter 2008) × body weight	-0.16	0.12	-1.33	0.184
Season (spring 2009) × body weight	-0.05	0.13	-0.39	0.698
Eye lens weight × sex (male)	630.45	245.61	2.57	0.011
Flea burden				
Intercept	-0.98	0.37	-2.65	0.009
Site (SQC)	-0.47	0.15	-3.16	0.002
Season (summer 2008)	0.85	0.21	4.01	<0.0001
Season (autumn 2008)	0.63	0.24	2.66	0.008
Season (winter 2008)	-0.54	0.32	-1.69	0.093
Season (spring 2009)	0.31	0.39	0.79	0.432
Body weight	0.04	0.01	4.26	<0.0001
Tick burden				
Intercept	-9.86	3.77	-2.61	0.009
Season (summer 2008)	-8.96	2.48	-3.62	<0.001
Season (autumn 2008)	2.51	4.35	0.58	0.564
Season (winter 2008)	-1.15	12.51	-0.09	0.927
Season (spring 2009)	1.88	6.90	0.27	0.785
Body weight	0.75	0.16	4.75	<0.0001
Eye lens weight	715.57	734.52	0.97	0.331
Season (summer 2008) × body weight	-0.27	0.07	-4.19	<0.0001
Season (autumn 2008) × body weight	-0.42	0.16	-2.67	0.008
Season (winter 2008) × body weight	-0.58	0.59	-0.98	0.329
Season (spring 2009) × body weight	-0.42	0.27	-1.55	0.122
Season (summer 2008) × eye lens weight	3450.54	577.21	5.98	<0.0001
Season (autumn 2008) × eye lens weight	1061.32	973.41	1.09	0.276
Season (winter 2008) × eye lens weight	2667.78	2722.31	0.98	0.328
Season (spring 2009) × eye lens weight	1489.39	1421.70	1.05	0.296
Body weight × eye lens weight	-97.97	27.82	-3.52	<0.0001

Appendix Table 2.3 GLMMs describing non-genetic factors associated with variation in the probability of infection by various parasites. ΔAIC = the change in the AIC if the single term is dropped. σ^2 = the variance attributable to a random effect. sd = standard deviation of σ^2.

Term	**Coefficient**	**s.e**	***z*-value**	***p*-value**	**ΔAIC**
Flea infection.					
Intercept	-1.47	0.56	-2.63	0.008	-
Site (SQC)	0.51	0.40	1.30	0.194	-
Season (summer 2008)	1.94	0.53	3.63	0.0002	4.3
Season (autumn 2008)	1.01	0.56	1.79	0.072	4.3
Season (winter 2008)	0.19	0.89	0.217	0.829	4.3
Sex (male)	0.722	0.334	2.16	0.031	-
Body weight	-0.003	0.019	-0.21	0.835	40.1
Site (SQC) × sex (male)	-1.05	0.438	-2.40	0.016	3.4
Random effect: *site*session*	σ^2 = 0.36 ; sd = 0.60				
Random effect: *individual*	σ^2 = 0.42 ; sd = 0.65				
Tick infection					
Intercept	-1.71	0.39	-4.40	>0.0001	-
Body weight	0.014	0.017	0.818	0.413	28.2
Random effect: *site*session*	σ^2 = 0.07 ; sd = 0.26				
Random effect: *individual*	σ^2 = 0.22 ; sd = 0.48				
Babesia microti *infection*					
Intercept	-1.58	0.39	-4.09	>0.0001	-
Site (SQC)	0.42	0.48	0.88	0.378	-
Season (summer 2008)	0.58	0.48	1.21	0.225	-
Season (autumn 2008)	0.68	0.44	1.53	0.125	-
Season (winter 2008)	0.13	0.77	0.17	0.869	-
Sex (male)	-0.70	0.34	-2.07	0.038	-
Body weight	0.03	0.02	1.74	0.082	19.6
Recapture (yes)	0.56	0.23	2.41	0.016	3.8
Site (SQC) × season (summer 2008)	-1.60	0.60	-2.68	0.007	3.5
Site (SQC) × season (autumn 2008)	-0.87	0.53	-1.63	0.102	3.5
Site (SQC) × season (winter 2008)	-1.66	1.09	-1.52	0.128	3.5
Site (SQC) × sex (male)	1.30	0.44	2.96	0.003	5.4
Random effect: *site*session*	σ^2 = 0.00 ; sd = 0.00				
Bartonella *infection*					
Intercept	0.13	0.26	0.51	0.612	-
Season (summer 2008)	0.80	0.32	2.47	0.013	12.5
Season (autumn 2008)	-0.45	0.34	-1.33	0.182	12.5
Season (winter 2008)	-2.51	0.74	-3.39	0.0006	12.5
Sex (male)	-0.22	0.21	-1.04	0.297	-
Body weight	-0.04	0.02	-1.73	0.083	-
Sex (male) × body weight	0.06	0.03	2.13	0.03	3.5
Random effect: *site*session*	σ^2 = 0.01 ; sd = 0.10				
Random effect: *individual*	σ^2 = 0.39 ; sd = 0.62				

Appendix Table 2.4 GLMMs describing non-genetic factors associated with variation in tick burden per individual. ΔAIC = the change in the AIC if the single term is dropped. σ^2 = the variance attributable to a random effect. sd = standard deviation of σ^2

Term	**Coefficient**	**s.e**	***t*-value**	**ΔAIC**
Intercept	-5.48	0.029	-191.88	-
Season (summer 2008)	3.05	0.03	106.62	-
Season (autumn 2008)	0.76	0.03	24.61	-
Season (winter 2008)	-0.57	0.06	-10.42	-
Sex (male)	0.80	0.02	49.69	-
Mature (yes)	1.76	0.03	63.56	-
Body weight	-0.15	0.00	-62.23	-
Season (summer 2008) × mature (yes)	-2.56	0.03	-83.50	7.7
Season (autumn 2008) × mature (yes)	-0.69	0.04	-19.51	7.7
Season (winter 2008) × mature (yes)	-10.70	17.57	-0.61	7.7
Sex (male) × body weight	0.14	0.00	95.46	9.7
Mature (yes) × body weight	0.15	0.00	64.59	5.2
Random effect: *site*session*	$\sigma^2 = 0.00$; sd = 0.06			
Random effect: *individual*	$\sigma^2 = 0.01$; sd = 0.08			

Appendix Table 2.5 Linear models describing non-genetic factors associated with variation in immune gene expression. Deletion testing led to the removal of all non-genetic terms in the models of *Il10* and *Irf5* expression.

Term	Coefficient	s.e	*t*-value	*p*-value
Foxp3 expression				
Intercept	-1.27	0.98	-1.29	0.198
Site (SQC)	-0.95	0.89	-1.07	0.286
Season (summer 2008)	1.01	0.46	2.21	0.025
Season (autumn 2008)	1.71	0.49	3.46	<0.001
Season (winter 2008)	1.12	0.50	2.26	0.026
Season (spring 2009)	1.30	0.62	2.10	0.037
Body weight	0.04	0.03	1.14	0.258
Eye lens weight	-23.36	221.55	-0.10	0.916
Sex (male)	-0.72	0.38	-1.90	0.059
Site (SQC) × body weight	-0.11	0.04	-2.34	0.020
Site (SQC) × eye lens weight	666.55	262.86	2.53	0.012
Site (SQC) × sex (male)	1.18	0.52	2.25	0.026
Gata3, 96 h PHA[a] expression				
Intercept	0.21	0.07	3.18	0.002
Season (summer 2008)	-0.17	0.10	-1.80	0.074
Season (autumn 2008)	0.16	0.11	1.49	0.136
Season (winter 2008)	-0.13	0.13	-1.01	0.316
Season (spring 2009)	-0.07	0.18	-0.38	0.706
Body weight	-0.003	0.002	-1.23	0.220
Season (summer 2008) × body weight	0.007	0.003	2.27	0.024
Season (autumn 2008) × body weight	-0.007	0.005	-1.43	0.154
Season (winter 2008) × body weight	0.004	0.006	0.63	0.529
Season (spring 2009) × body weight	0.003	0.009	0.31	0.758
Ifng expression				
Intercept	-1.59	0.64	-2.48	0.014
Site (SQC)	0.63	0.27	2.34	0.020
Season (summer 2008)	0.05	0.40	0.11	0.910
Season (autumn 2008)	-0.75	0.39	-1.95	0.053
Season (winter 2008)	1.70	0.40	4.27	<0.001
Season (spring 2009)	1.62	0.56	2.88	0.005
Body weight	0.04	0.02	1.99	0.049
Il1b expression				
Intercept	-2.27	1.45	-1.57	0.119
Season (summer 2008)	-1.07	0.37	-2.86	0.005
Season (autumn 2008)	-1.38	0.40	-3.45	<0.001
Season (winter 2008)	0.17	0.42	0.40	0.689
Season (spring 2009)	-0.49	0.55	-0.89	0.37
Body weight	0.13	0.06	2.34	0.020
Eye lens weight	651.67	314.19	2.07	0.039
Body weight × eye lens weight	-29.83	11.93	-2.50	0.013

[a] *Gata3* expression measured at 96 h from splenocyte cultures stimulated with the mitogen PHA

Appendix Table 2.5 continued

Term	Coefficient	s.e	*t*-value	*p*-value
Il2 expression				
Intercept	-0.18	0.30	-0.60	0.550
Season (summer 2008)	-0.25	0.44	-0.56	0.576
Season (autumn 2008)	-0.18	0.39	-0.47	0.638
Season (winter 2008)	0.65	0.38	1.72	0.087
Season (spring 2009)	1.56	0.59	2.62	0.009
Il10 expression				
Null	-	-	-	-
Irf5 expression				
Null	-	-	-	-
Tbx21 expression				
Intercept	-0.65	0.30	-2.19	0.030
Site (SQC)	0.45	0.24	1.91	0.058
Season (summer 2008)	-0.31	0.36	0.88	0.379
Season (autumn 2008)	-1.02	0.32	-3.22	0.005
Season (winter 2008)	1.23	0.31	3.96	<0.001
Season (spring 2009)	1.50	0.48	3.10	0.002
Sex (male)	0.43	0.21	1.20	0.048
Tgfb1 expression				
Intercept	-0.55	0.31	-1.79	0.076
Season (summer 2008)	-0.45	0.52	-0.86	0.39
Season (autumn 2008)	0.28	0.43	0.65	0.51
Season (winter 2008)	1.07	0.39	2.77	0.006
Season (spring 2009)	1.15	0.84	1.37	0.174
Sex (male)	0.78	0.41	1.87	0.063
Season (summer 2008) × sex (male)	-0.39	0.66	-0.59	0.555
Season (autumn 2008) × sex (male)	-0.07	0.56	-0.13	0.900
Season (winter 2008) × sex (male)	-1.38	0.52	-2.63	0.010
Season (spring 2009) × sex (male)	-0.59	1.10	-0.53	0.592
PC^{I} expression [b]				
Intercept	-0.02	0.26	-0.07	0.943
Season (summer 2008)	-0.83	0.41	-2.01	0.046
Season (autumn 2008)	-1.21	0.37	-3.28	0.001
Season (winter 2008)	1.03	0.34	3.06	0.003
Season (spring 2009)	1.23	0.64	1.92	0.057
PC^{AI} expression [c]				
Intercept	-4.96	1.96	-2.53	0.013
Season (summer 2008)	0.42	0.78	0.54	0.590
Season (autumn 2008)	0.80	0.70	1.15	0.252
Season (winter 2008)	1.95	0.68	2.85	0.005
Season (spring 2009)	1.84	1.09	1.68	0.096
Body weight	0.14	0.07	1.86	0.064
Eye lens weight	1020.76	395.04	2.58	0.011

[b] Grouped (PCA) pro-inflammatory response

[c] Grouped (PCA) anti-inflammatory response

Appendix Table 2.5 continued

Term	Coefficient	s.e	*t*-value	*p*-value
PC^{AI} expression, cont.				
Sex (male)	1.16	0.60	1.96	0.053
Season (summer 2008) × sex (male)				
Season (autumn 2008) × sex (male)				
Season (winter 2008) × sex (male)				
Season (spring 2009) × sex (male)				
Body weight × eye lens weight				

Appendix Table 2.6 GLM describing non-genetic factors associated with unstimulated *Gata3* expression measured at 96 h.

Term	Coefficient	s.e	*t*-value	*p*-value
Intercept	1.11	1.88	0.59	0.554
Season (summer 2008)	-0.31	0.81	-0.38	0.703
Season (autumn 2008)	0.54	0.88	0.62	0.538
Season (winter 2008)	-1.71	1.11	-1.53	0.125
Season (spring 2009)	-0.55	1.29	-0.43	0.670
Eye lens weight	-874.72	82.77	-1.81	0.070
Sex (male)	-3.73	1.74	-2.13	0.032
Eye lens weight × sex (male)	1043.78	520.12	2.00	0.045